From the Department of Medical Biochemistry and Biophysics
Karolinska Institutet, Sweden, 2015

Structural and Functional Studies of Nutrient and Drug Uptake Systems

Karolinska Institutet

Stockholm 2015

Published by Karolinska Institutet.
Printed by US-AB

ISBN 978-91-7549-979-6

Structural and functional studies of nutrient and drug uptake systems

THESIS FOR DOCTORAL DEGREE (Ph.D.)

By

Fatma Guettou

Principal Supervisor:
Pär Nordlund
Karolinska Institutet
Department of Medical Biochemistry and Biophysics

Co-supervisor:
Christian Löw
EMBL Hamburg c/o DESY

Opponent:
David Drew
Stockholm University
Department of Biochemistry and Biophysics

Examination Board:
Lena Mäler
Stockholm University
Department of Biochemistry and Biophysics

Christina Divne
Karolinska Institutet
Department of Medical Biochemistry and Biophysics

Doreen Dobritzsch
Uppsala University
Department of Chemistry

Till min familj

ABSTRACT

Orally ingested nutrients and drugs are selectively absorbed from our small intestines into the bloodstream through various membrane-integrated transporters. The present study focuses mainly on a specific absorption route, namely the proton dependent oligopeptide transporters (POTs). These transporter systems belong to the major facilitator superfamily (MFS) and are secondary active transporters. The aim of this thesis was to study the structure and mechanism of POTs from prokaryotic organisms. The project was divided in two phases. During the first phase, a high-throughput method was developed for rapid screening of integral membrane proteins (IMP) to identify suitable targets, constructs, and production conditions for structural studies (paper I). During the second phase of the project, X-ray crystal structures of the prokaryotic peptide transporter ($PepT_{So2}$) from the organism *Shewanella oneidensis* in complex with four different substrates were determined (paper II-III). The structures revealed the overall conformational state of the protein as well as the architecture of the substrate-binding site. The protein was captured in an inward open conformation where the substrate-binding site was accessible to the cytoplasm but not to the periplasm. The bound peptides adopted extended lateral conformations with their N-termini interacting with a conserved polar pocket while their C-termini were in close proximity to a positively charged pocket. The results presented in papers I-III provide novel structural and mechanistic insights into prokaryotic peptide transporters. Interestingly, the binding site residues are highly conserved in the human peptide transporter homolog, PepT1. Hence, these results not only increase our understanding regarding prokaryotic peptide transporters but also shed light on the human homologs. Furthermore, results presented in this work may assist in design of pharmacologically active compounds into substrates of the human peptide transporter, creating orally administrated drugs.

LIST OF SCIENTIFIC PAPERS

The following papers are included in this thesis

I. High-throughput analytical gel filtration screening of integral membrane proteins for structural studies.
Löw C, Moberg P, Quistgaard EM, Hedrén M, Guettou F, Frauenfeld J, Haneskog L and Nordlund P.
Biochim. Biophys. Acta. 1830(6), 3497-508 (2013).

II. Structural insights into substrate recognition in proton-dependent oligopeptide transporters.
Guettou F, Quistgaard EM, Trésaugues L, Moberg P, Jegerschöld C, Zhu L, Jong AJ, Nordlund P and Löw C.
EMBO Rep. 14(9), 804-10 (2013).

III. Selectivity mechanism of a bacterial homolog of the human drug-peptide transporters PepT1 and PepT2.
Guettou F, Quistgaard EM, Raba M, Moberg P, Löw C and Nordlund P.
Nat. Struct. Mol. Biol. 21(8), 728-31 (2014).

The author contributed to the following publications, not included in this thesis

I. Structural and biophysical characterization of the cytoplasmic domains of human BAP29 and BAP31.
Quistgaard EM, Löw C, Moberg P, Guettou F, Maddi K, Nordlund P.
PLoS One. 13(8), e71111 (2013)

II. A multi-functional nanoparticle system based on a small human protein.
Frauenfeld F, Löving R, Zhang Y, Zhu L, Jegerschöld C, Guettou F, Moberg P, Löw C, Nyström N, Garoff H and Nordlund P.
Nat. Nanotechnol.
Submitted

III. Mechanism of the major facilitators – a structure based review.
Quistgaard EM, Löw C, Guettou F, and Nordlund P.
Nat. Commun.
Submitted

TABLE OF CONTENTS

Abbreviations

ABC-transporter	ATP-binding cassette transporter
CMC	Critical micelle concentration
IMP	Integral membrane protein
DDM	n-Dodecyl β-D-maltoside
His-tag	Histidine-tag
IMAC	Immobilized metal affinity chromatography
IM	Inner membrane
IPTG	Isopropyl β-D-1-thiogalactopyranoside
MAD	Multiple anomalous dispersion
MIR	Multiple Isomorphous replacement
MR	Molecular replacement
MFS	Major facilitator superfamily
OM	Outer membrane
pET vector	Plasmid for expression by T7 RNA polymerase
POT	Proton dependent oligopeptide transporter
RNAP	RNA polymerase
SEC	Size exclusion chromatography
TM	Transmembrane
TMD	Transmembrane domain
XFEL	X-ray free electron laser

Amino acids

Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartate	Asp	D
Cysteine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

1 Introduction

The work presented in this thesis focuses on structural and functional understanding of proteins involved in nutrition and drug uptake. The first chapter offers an overview of basic concepts and principles relevant for the studies performed. The second and third chapters describe crystallography and protein structure determination methods and the forth chapter focuses on channel and transporter proteins.

1.1 The central dogma of molecular biology

The central dogma is a well-established principle explaining the connection between DNA, RNA, and proteins. It was first stated by Francis Crick in 1956 [1] at a time when little evidence supported these ideas, hence the name dogma. It comprises three major processes for the transfer of information between DNA, RNA, and protein. The first process is replication: parental DNA is replicated to form a copy DNA by a set of enzymes where the main actor is DNA-polymerase. The second process is transcription: DNA is transcribed into RNA by RNA-polymerase. The third process is translation: RNA is translated by the ribosome into a polypeptide chain of amino acids resulting in a protein (Figure 1).

Besides these three general processes, two additional routes exist: RNA can be replicated by the enzyme RNA dependent RNA polymerase and DNA can be generated from RNA by the retrovirus enzyme reverse transcriptase. Nevertheless information can never be reversed from a protein back to nucleic acids [2].

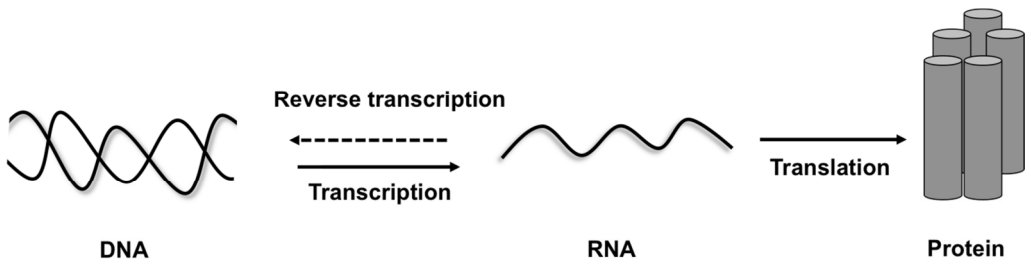

Figure 1. An illustration of the central dogma. Double stranded DNA is transcribed into a single stranded RNA which is then translated into a protein.

1.2 Proteins

The word protein was first mentioned 1838 by the Swedish chemist, Jöns Jacob Berzelius, in one of his letters to the Dutch chemist, Gerardus Johannes Mulder [3].

"Le nom protéine que je vous propose pour l'oxyde organique de la fibrine et de l'albumine, je voulais le dériver de proteios parce qu'il parait être la substance primitive ou principale de la nutrition animale"

Today we know that proteins are essential biological molecules that constitute the largest fraction of a cell where they are responsible for a vast array of functions. The structure of a protein can be divided in four levels: primary, secondary, tertiary and quaternary structure (Figure 2). The primary structure describes the sequence of amino acids that are linked in a specific order through covalent bonds. The linear primary structure folds into secondary structure elements, typically α-helices, β-sheets and turns. α-helices are right-handed spiral shaped structures while β-sheets are planar strands, both structures are stabilized by multiple hydrogen bonds [4]. The next structural level is the tertiary structure, this is usually the structure referred to when describing a protein. It is stabilized by multiple hydrogen bonds together with salt-bridges, disulfide bonds and hydrophobic interactions. The fourth structural level, the quaternary structure, describes a large assembly of several proteins interacting to form larger oligomer arrangements. There are two types of oligomeric structures: homo- and hetero oligomers [5]. As the names indicate, homo-oligomers are protein complexes composed of identical units held together by non-covalent interactions. In contrast, hetero-oligomers are protein complexes composed of different units. For example, hemoglobin is composed of four identical subunits and is thus referred to as a homo-tetramer. Eukaryotic ATP-binding cassette transporters (ABC transporters) are heteromeric complexes composed of an ATPase subunit and a transmembrane subunit. The different subunits in an oligomeric assembly may possess either independent or cooperative activity. Oligomeric protein complexes are abundant in nature, they offer advantageous features enabling the protein to reach higher orders of complexity including efficient pathways for functional control.

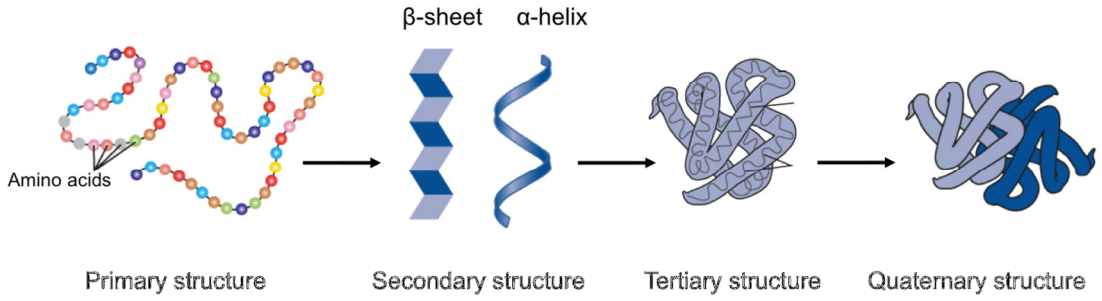

Figure 2. An illustration of the four protein structure hierarchies. The primary structure describes the linear sequence of amino acids. The secondary structure elements are mainly formed by α-helices and β-sheets. The secondary structure elements are organized into a tertiary structure giving the native shape of a protein. The quaternary structure is a complex composed of several proteins that are arranged either as a homo-oligomer or a hetero-oligomer. Figure adapted from the National Human Genome Research Institute.

1.2.1 Membrane proteins

Membrane proteins are hydrophobic molecules that are embedded in the complex lipid environment of the cell membrane. A typical cell membrane holds a broad set of different membrane proteins, all specialized in a particular function. They enable the cell to carry out various actions including the import of metabolites and export of waste, transmission of chemical signals and storage of chemical energy. Some membrane proteins are spanning the cell membrane, whereas others are bound to the membrane surface. Based on these characteristics, they are usually classified in two large groups: peripheral membrane proteins and integral membrane proteins (IMPs) (Figure 3).

IMPs are fully integrated into the cell membrane by spanning it once or several times. Their amino acid composition is mostly hydrophobic thus enabling interaction with the hydrophobic cell membrane (Figure 3). Many IMPs create sluices through the membrane for the transport of various substances, this ability is based on their amphipathic secondary structures where the amino acid side chains alternate between polar and nonpolar property. The nonpolar side chains interact with the membrane acyl chains while the polar side chains form the inside of the pore. In this way, polar substances can be transported though the otherwise hydrophobic membrane [6].

Peripheral membrane proteins are loosely attached to the membrane and they do not interact with the hydrophobic core of the membrane as extensively as the IMPs (Figure 3). Instead, they attach in a reversible manner to the membrane. The reversible nature enables

them to shuttle between the membrane and the cytosol allowing them to be involved in tasks such as cell signalling [7].

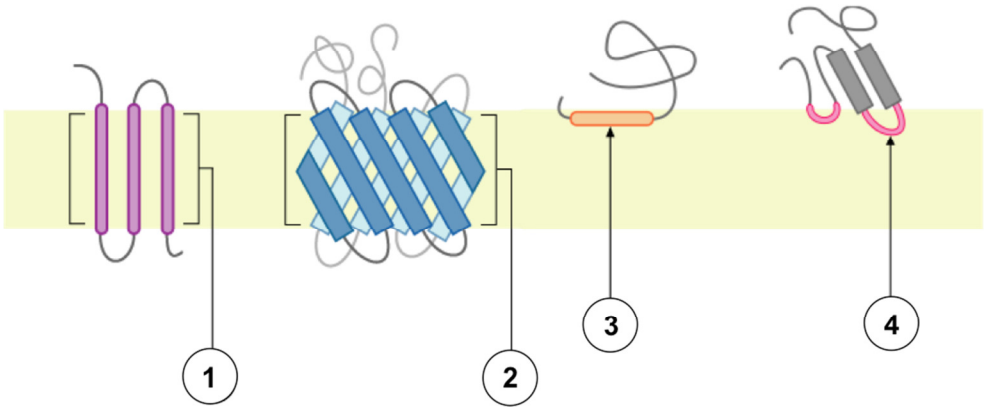

Figure 3. Schematic representation of different types of membrane proteins. 1-2 shows IMPs that span the cell membrane as alpha helices or beta-sheets respectively. 3-4 shows two ways for peripheral membrane proteins to interact with the cell membrane.

1.3 The E. coli cell membrane

Escherichia coli (*E. coli*) is the most studied prokaryotic model system from where much of our knowledge about genetics, biochemistry and biophysics is based. A typical *E. coli* cell is rod shaped with the dimensions of 2 - 4 µm in length and 1 - 2 µm in width [8]. It is part of our normal gut flora and it can be found in the lower intestines of warm-blooded organisms. An *E. coli* cell is surrounded by a protective membrane made up by a lipid bilayer. Lipids are amphiphilic molecules consisting of a hydrophilic head group and two hydrophobic tails. In a polar solution, lipids self-organize into bilayers by exposing the hydrophilic head groups to the polar solutions while hiding their hydrophobic tails inside the bilayers. The lipid bilayers can be further divided into an inner and an outer leaflet (Figure 4). There are numerous types of lipids, all with different chemical properties of the head groups and the tail structures. In the following section, the different compartments of an *E. coli* cell membrane: the outer membrane (OM), the periplasm, and the inner membrane (IM) will be described (Figure 4).

The two leaflets of the *E. coli* OM are asymmetrical with regards to their lipid constitution. The outer leaflet is composed mainly of lipopolysaccharides; they maintain the rigidity of the cell, contribute to the negative charge of the membrane and are responsible for the immune response initiation in mammals when exposed to pathogenic *E. coli* stains [8, 9]. The inner leaflet of the OM, on the other hand, is composed of a mixture of three glycerol-

phospholipids: the zwitterionic glycerol phosphatidyl ethanolamine (PE), the anionic phosphatidylglycerol (PG), and the anionic cardiolipin (CL) [9]. IMPs located in the OM are predominantly composed of β-strands forming β -barrels. Porins are an example of OM β-barrel proteins responsible for the passage of a wide range of molecules over the OM.

The periplasm is a compartment sandwiched between the OM and the IM. It encloses a strong peptidoglycan mesh layer that gives the cell its shape and structure.

The two leaflets of the IM consist mainly of glycerol-phospholipids [9]. The IM harbours a range of membrane proteins mainly with an α-helical structure, proton dependent oligopeptide transporters are examples of α-helical IM proteins.

It should be mentioned that lipids also possess non-membrane related functions. They are for example involved in biochemical signalling where they act as secondary messengers. One of the examples is the lipid triacylglycerol (DAG) which is involved in the activation of protein kinase C responsible for substrate phosphorylation [10].

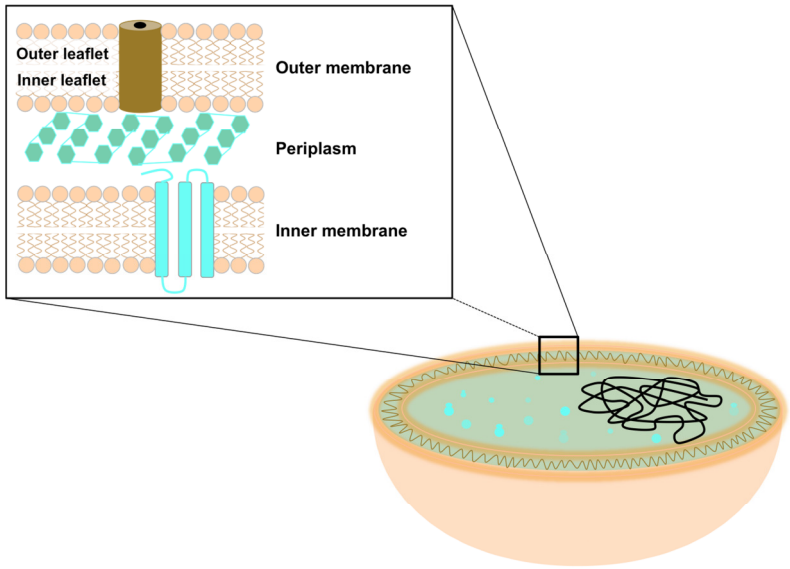

Figure 4. Schematic illustration of an *E. coli* bacterium. The cell is cut in half to visualize the content of the cytosol and cell membrane. The OM is exposed to the outside environment and contains β-barrel proteins responsible for the transport of molecules. The IM contains α-helical IMPs responsible for the transport of molecules from and to the cytosol. The periplasm is a rigid structure sandwiched between the OM and the IM.

1.4 Recombinant protein production in E. coli

To pursue structural studies on IMPs, large quantities of pure and homogenous protein is required. However, the natural abundance of these proteins is often too low to accumulate. Therefore, recombinant production of IMPs in *E. coli* has proven to be a robust strategy. The *E. coli* cell is in fact used as a protein factory, offering in optimal cases an easy-to-handle way of producing adequate quantities of proteins at low cost. Another advantage is the possibility to screen for optimal protein production strategies by varying for example the *E. coli* strains and expression conditions. Nevertheless, there are some bottlenecks in the *E. coli* IMP production process. Some of the most common issues are low expression yields, other more serious issues are cell toxicity or formation of aggregated proteins called inclusion bodies. The reason for such effects is the enormous stress imposed on the *E. coli* transcription and translation systems during overexpression of a particular protein [11].

1.4.1 BL21(DE3)/pET protein production system

BL21(DE3) is an *E. coli* strain specially generated in 1986 by Studier *et al.* with the aim of improving protein production [12]. Two main modifications are present in this strain. The first modification is the introduction of a lambda bacteriophage DNA segment (DE3) carrying the *lacI* gene, the *lacUV5* promoter, the *lacZ* gene followed by the T7 RNA polymerase (RNAP) gene, all located at the *E. coli* chromosomal DNA. The second modification is the introduction of an extra-chromosomal circular DNA vector, pET-vector (plasmid for expression by T7 RNA polymerase), containing a T7 promoter region upstream of the multiple cloning site where the target protein gene is inserted (Figure 5). This expression system takes advantage of the high selectivity of T7 RNAP for the T7 promoter. The target protein gene can therefore only be transcribed by the T7 RNAP. Nevertheless, the transcription of T7 RNAP is governed by the *lac*UV5 promoter, a mutated version of the *lac* promoter. Under normal circumstances, a *lac* repressor is bound to the *lac* operon, thus inhibiting the *E. coli* RNAP from transcribing T7 RNAP. Only when the lactose mimic Isopropyl β-D-1-thiogalactopyranoside (IPTG) is added to the cells, the repressor is released and T7 RNAP is transcribed. Subsequently, the T7 RNAP transcribes the target protein gene located in the pET vector.

The BL21(DE3) strain was further improved by knocking out two proteases, the outer membrane protease OmpT and the cytoplasmic protease Lon [13,14]. Both these proteases were

known for having a negative effect on the target protein during expression and purification. Taken together, the combination of the pET vector, the *lac*UV5 promoter and the knock-out of OmpT and Lon make the *E. coli* strain BL21(DE3) a highly powerful and appreciated tool for heterologous protein production.

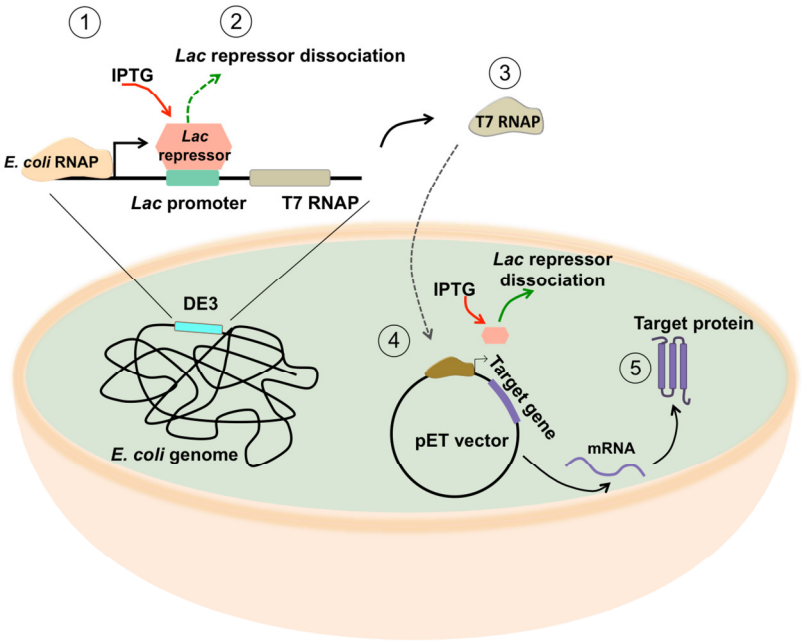

Figure 5. An illustration of the BL21(DE3) expression system. Under normal circumstances a *lac* repressor is bound to the *lac* operon inhibiting production of T7 RNAP. Addition of IPTG (1) results in the dissociation of the *lac* repressor (2) enabling the production of T7 RNAP (3). The produced T7 RNAP recognizes its T7 promoter on the pET vector and is thus responsible for the transcription of the target gene (4). The resulting target mRNA is then translated by the ribosome.

1.4.2 Detergents

Detergents are typically used for the isolation and solubilization of IMPs from the cell membrane since they offer a lipid mimetic environment for hydrophobic proteins. The general structure of detergents is similar to lipids - a polar head group fused to a nonpolar hydrophobic acyl chain tail. However there are two major differences between them: first, detergents consist of only one acyl chain tail structure while lipids consist of two, second, detergents used for IMP solubilization are cone shaped enabling them to form micelle structures while bilayer-forming lipids have a cylindrical overall shape (Figure 6 A-B) [15].

When exposed to aqueous solution, detergent micelles form spontaneously only if the detergent concentration is above the critical micelle concentration (CMC). Exposing the cell membrane to detergents at a concentration above their CMC will result in protein:detergent complexes where the detergent tail structures interact with the hydrophobic parts of the proteins while the hydrophilic detergent head group keeps the complex soluble (Figure 6 C).

Detergents are often classified into three groups based on the nature of their polar head groups: nonionic, zwitterionic or ionic. Nonionic detergents consist of a hydrophilic head group such as maltose and are generally mild on IMPs as they do not disrupt any structural features, N-Dodecyl-β-D-maltopyranoside (DDM) is a widely used nonionic detergent. Zwitterionic detergents possess a net zero charge just like nonionic detergents, yet they are less mild as they might disrupt protein interactions. A typical zwitterionic detergent is lauryldimethylamine oxide (LDAO). Ionic detergents are efficient in solubilizing membrane proteins but they are also the least mild type of detergents since they usually disrupt protein interactions. Sodium dodecyl sulfate (SDS) is an anionic detergent used for protein denaturation.

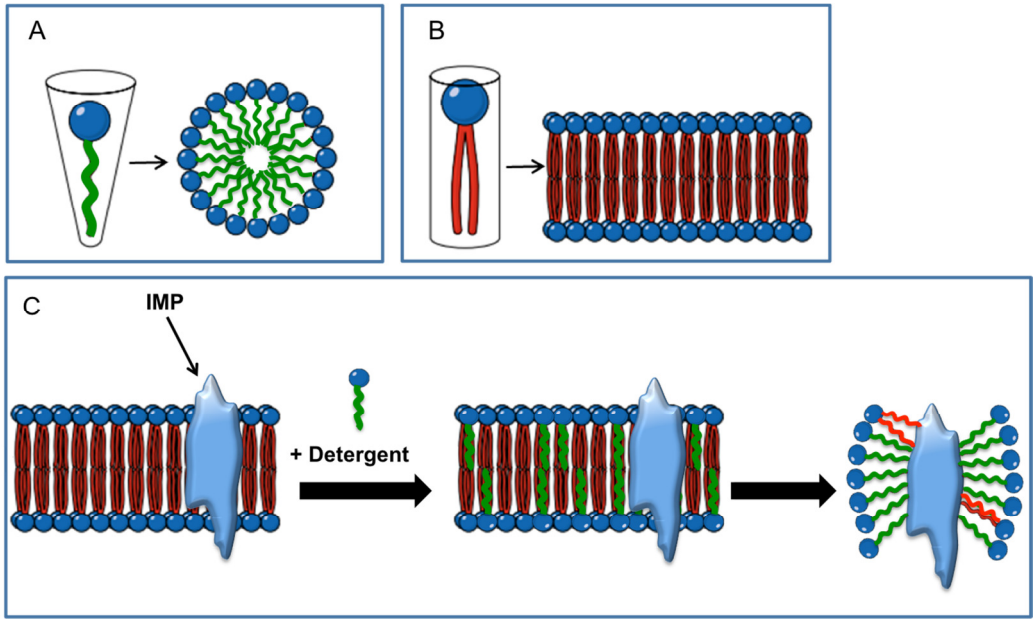

Figure 6. Detergents and lipids. A) The overall cone shape of detergents enables them to form micelles. B) The overall cylindrical shape of lipids enabling them to form bilayers. C) Illustration of IMP solubilization: the detergent destabilizes the lipid bilayer and extracts the protein from the membrane. The final protein-detergent complex might also contain lipids. Figure adapted from [16]

1.4.3 Purification of Integral membrane proteins

The aim of protein purification is to isolate a specific target protein from a complex mixture by successive purification until a state of homogeneity is reached. Common purification strategies include steps such as centrifugation, affinity chromatography and size-exclusion chromatography where protein characteristics such as size, chemical features and affinity tags are exploited for separation.

The first step in a typical IMP purification procedure is the cell lysis where cells are broken by chemical or mechanical means, repeated freeze-thawing cycles, sonication or homogenization. Lysed cells are then subjected to low speed centrifugation for the isolation of cell debris followed by high-speed centrifugation, for the isolation of the cell membrane fraction [17].

Subsequently, IMPs need to be extracted from the membrane fraction. This is usually accomplished by solubilizing the membrane fraction in detergents providing a membrane-mimetic environment for IMPs. Finding a suitable detergent for a specific IMP is an intricate process, there are no general rules to follow hence the choice of detergent has to be optimized experimentally for each target protein. Choosing a suboptimal detergent might result in inefficient or heterogeneous solubilization, loss of protein function and stability or in the worst case, protein precipitation. Therefore, detergent screening is usually performed such as described in Paper I [18].

Detergent solubilized IMPs are commonly purified by affinity chromatography, a widely used technique based on the interaction between two molecules such as antigen and antibody. Immobilizing one interaction partner in the stationary phase of a column enables the binding and enrichment of the other interaction partner. Immobilized metal ion affinity chromatography (IMAC) is based on the interaction between histidine residues and divalent ions such as nickel. For this technique to work it is required that the target protein is fused to a poly histidine tag (His-tag) while the column stationary phase contains nickel. When running an IMAC experiment, His-tagged protein will be retained by the nickel on the column while other proteins will be washed away. The His-tagged protein is then released from the column by various methods such as, pH change or addition of a competitive molecule such as imidazole (Figure 7 A) [19, 20].

As a final purification step, size exclusion chromatography (SEC) is applied after IMAC purification. Here, proteins are separated on a column containing an inert porous matrix of spherical particles that allow separation of proteins based to their size. Unlike IMAC, SEC is

a nonbinding chromatographic method where proteins are eluted isocratically, without the need of a special buffer for elution. Large molecules exit the column before smaller molecules because small molecules get trapped inside the pores of the matrix. Protein elution is typically monitored by UV absorption at 280 nm while salt elution is monitored by its conductivity [21]. A SEC run is displayed as a chromatogram where a monodisperse elution profile is indicative for a homogeneous protein sample (Figure 7 B).

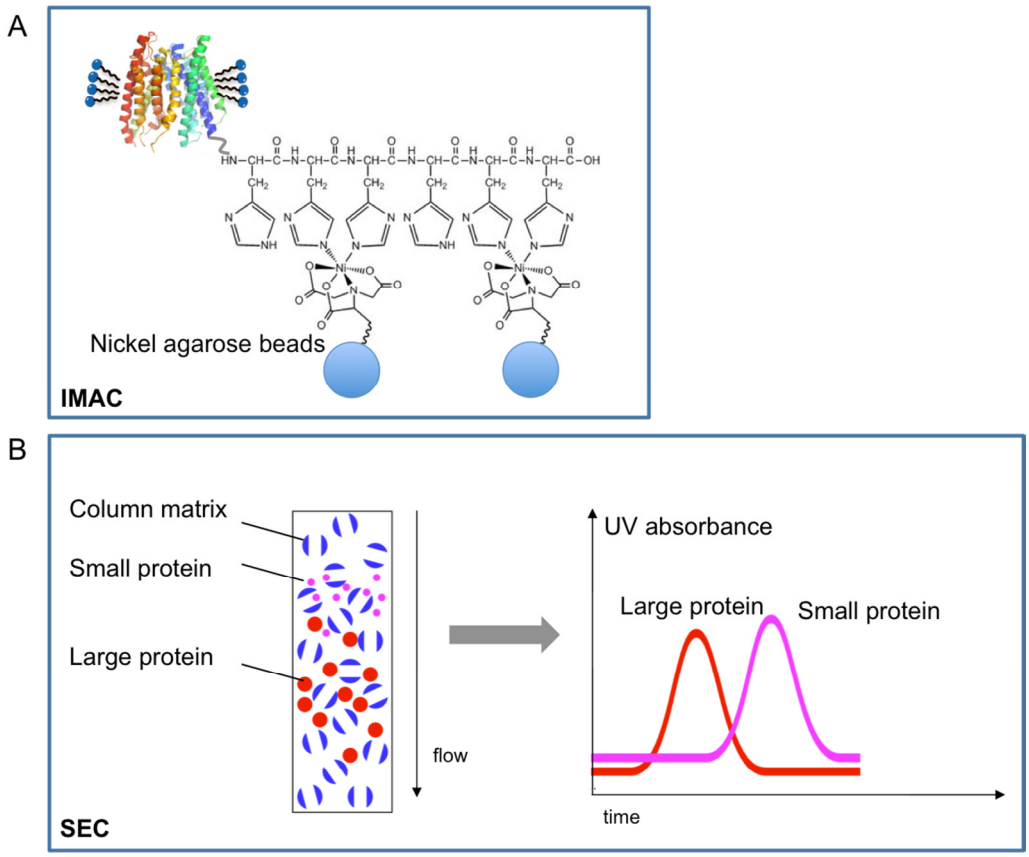

Figure 7. The purification principles of IMAC and SEC. A) The His-tagged recombinant protein binds specifically to the nickel agarose stationary phase while untagged proteins are washed away. B) The matrix of a SEC column separates protein according to their size. The chromatogram shows a shorter retention time for large proteins (red) compared to small proteins (purple). Figure adapted from [21].

1.5 Thermal shift assay

Natively folded proteins denature if subjected to extreme conditions such as high or low pH, elevated temperature or high salt concentrations. Thermal shift assays are based on the concept of destabilization and unfolding of proteins when subjected to elevated temperatures [22]. The protein stability is represented by the melting temperature (T_m), a specific temperature where 50% of the protein in a sample is unfolded. In general, T_m is altered when a ligand or substrate binds to the protein, normally resulting in an increased intrinsic thermal stability [23]. These features were described already in the 1950s by Koshland, Linderstrom-Lang and Schellman [24], and are utilized when studying protein-ligand interactions. The general approach is to heat up protein samples with and without a potential ligand in a controlled manner while monitoring the folding state as a function of temperature (Figure 8). Usually, a temperature shift greater than 2 °C is considered as a significant shift for a binding event.

There are various methods used for detecting the thermal stability of proteins. The Differential Static Light Scattering method is one of them [25]. It is a commonly used method based on the assumption that once a protein unfolds it immediately aggregates. Thus, monitoring the aggregation process (T_{agg}) through light scattering is an indirect way of measuring protein unfolding [26]. This type of analysis is ideal for IMPs since no additional labelling or couplings of fluorescent dyes is needed and there is no interference with the detergents.

The unfolding process may also be monitored by more conventional methods such as Sodium Dodecyl Sulphate Polyacrylamide Gel Electrophoresis (SDS-PAGE) or analytical gelfiltration analysis. Here, the sample is heated up with and without a ligand in a PCR thermal cycler to various temperatures followed by centrifugation to remove aggregated material. The concentration of remaining protein is then analysed by quantifying it on SDS-PAGE or by peak intensity from an analytical gel filtration run.

Another method, developed by Raymond Stevens *et al.* [27], is based on monitoring the thermal stability of proteins by a thiol-specific fluorochrome (N-[4-(7-diethylamino-4-methyl-3-coumarinyl) phenyl] maleimide) that binds to core thiol residues upon protein thermal unfolding. This method is highly sensitive and compatible with detergent imbedded IMPs, however the major drawback is the requirement of free thiol residues in the protein core.

Thermal shift assays have a very broad application area where protein stability can be analysed in various conditions. Examples of screening experiments include: finding an optimal buffer or detergent conditions and screening for ligands, substrates and binding

partners. Furthermore, a stabilizing molecule may be used for crystallization trials where it can either promote better quality crystals or, as performed in Paper II-III, identify a potential ligand for co-crystallization studies.

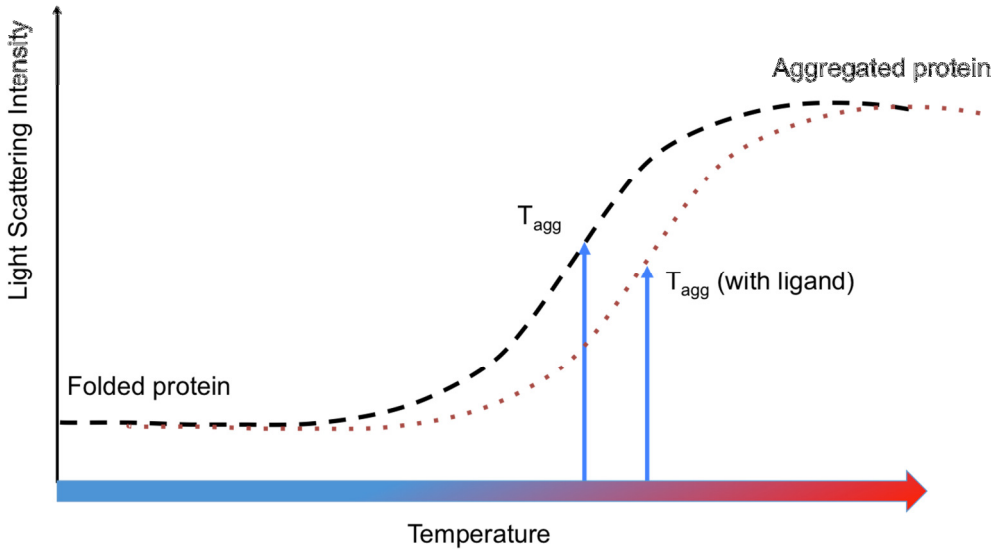

Figure 8. A typical thermal aggregation curve showing thermal stabilization effect upon ligand binding.

2 X-ray crystallography

The German physicist Wilhelm Conrad Röntgen was the first to report on X-rays in a paper published 1895 [28]. He discussed the properties of X-rays such as their ability to penetrate matter, to propagate in straight lines and to interact with photo-paper. Shortly after this discovery, X-rays were used to study the atomic constitution and bond interactions in salt crystals and diamonds. Already during this time it was known that proteins could form crystals. However, the techniques available were not sophisticated enough to elucidate complex protein structures. The first protein structure to be determined was a 2 Å structure of myoglobin that took 23 years for John Kendrew and Max Perutz to fully solve [29,30]. The first membrane protein structure to be determined was a subunit from the photosynthetic reaction centre, solved in 1985 by Deisenhofer and colleagues [31]. These first crystallographers were true pioneers, the importance of structural information to understand the molecular mechanism of biological system is therefore reflected by the large number of Nobel prizes that have been awarded for crystallographic work.

Apart from X-ray crystallography, there are diverse methods that can be applied for structural studies. Some of the most popular current methods are nuclear magnetic resonance (NMR) and single particle cryo-electron microscopy (Cryo-EM). NMR is used for the study of protein structure, dynamics, and kinetics, however, there are limitations regarding the size of the protein where smaller proteins are preferred [32]. Although order parameters (eg. B-factors) and extent of disordered/flexibility can be obtained from crystallographic structures, NMR can give additional valuable information on protein dynamics which is not directly obtained in crystal structures. Cryo-EM is a promising technique for challenging proteins including IMPs. Based on new computation approaches and detectors, IMP structures can, at least in special cases, be resolved at atomic resolution without the need to crystallize or solve the phases of target protein. However, protein size is also a significant limitation for Cryo-EM where larger proteins are preferred [33].

2.1 Why do we need crystals?

Protein crystals are fundamental in the course of macromolecular X-ray crystallography, the reason being that diffraction from a single molecule is too weak to be detected. Crystals are therefore suitable because of their periodic network of repeating protein arrays leading to an amplification of the diffracted X-ray signal. The smallest repeating units of a crystal are called unit cells and they are arranged according to a space group symmetry describing the position of each protein in the crystal [34] (Figure 9). One of the most frequently obtained space groups for membrane proteins is the $P2_12_12_1$, consisting of a two-fold symmetry combined with a screw axis.

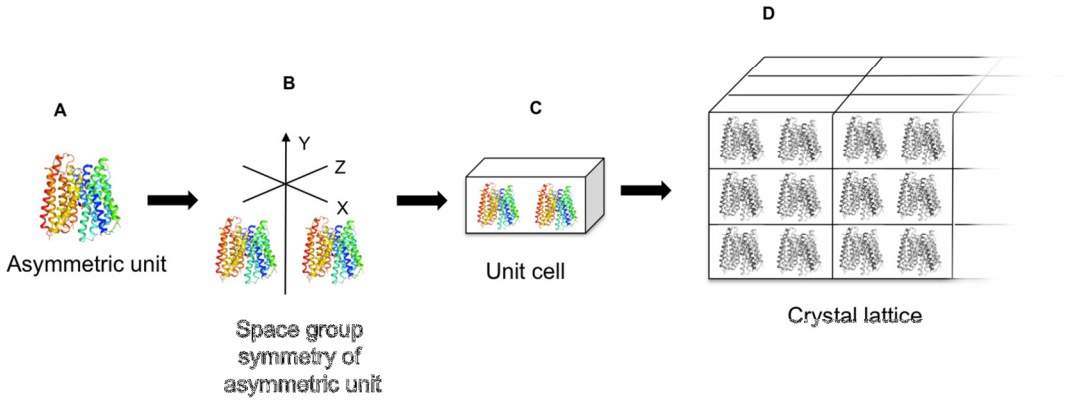

Figure 9. A) The asymmetric unit is the smallest part that when arranged by the space group symmetry (B) can produce the unit cell (C). The unit cell is the simplest repeating unit in a crystal (D).

Proteins are large, dynamic and flexible molecules whose native state rarely is a crystalline state which makes their crystallization a challenging process. Furthermore, protein crystals also need to be of a certain size and quality for high resolution X-ray diffraction. Moreover, there is no identified method for the prediction of an appropriate crystallization condition for a given protein. Indeed growing well-diffracting crystals has proved to be one of the major bottlenecks in the process of structure determination. Often, empirical screening of many crystallization conditions is necessary, where buffers, salts and organic solvents are varied, before a suitable condition is found. The crystallization process is thus a systematic search for parameters promoting crystal formation and due to the large number of possible parameters the search is often labour intensive and can take several years to achieve.

2.1.1 Protein crystallization – Vapour diffusion

Vapour diffusion is the most commonly used method for crystallizing a protein. An isolated drop of concentrated homogenous protein is mixed with a precipitant solution and kept either as a sitting or hanging drop next to a reservoir containing higher precipitant concentration. Initially, this system is not in equilibrium but will eventually move towards equilibrium by vapour diffusion from the protein drop to the reservoir solution. Consequently, the protein concentration will gradually increase as vapour is escaping from the protein drop. If the conditions are optimal, the protein will pass through the undersaturated zone and the metastable zone and end up in the nucleation zone (Figure 10). This is a zone where the attractive forces between proteins are maximized, an ideal environment for a protein nucleus to grow. As the nucleus grows, the free protein concentration in the drop will decrease and the system will move to the metastable zone, where a crystal can slowly grow to maturity [35].

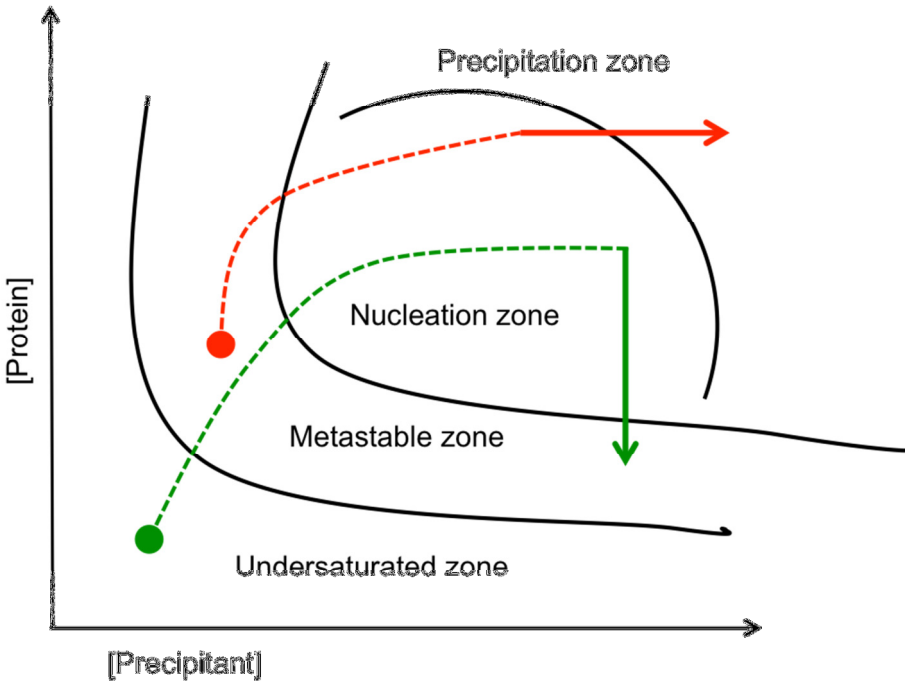

Figure 10. The phase diagram for a crystallization event. The diagram consists of four zones where a successful crystallization event follows the green line. A suboptimal crystallization condition might result in precipitation (red line). This simplified diagram is controlled by the precipitant and protein concentration.

2.1.2 Crystallization optimization strategies

The hydrophobic nature and the low intrinsic stability of IMPs make them very difficult to crystallize. Some of the screening strategies, used for membrane protein crystallization, are discussed below.

Solubilizing IMPs from the cell membrane with detergents result in a protein-detergent complex where the detergents form a belt-like structure around the hydrophobic parts of the protein (Figure 6 C). The detergent not only keeps the protein soluble but may also shield possible crystal contacts. As a rule of thumb: longer chain detergents (low CMC) are more efficient in the solubilization process but also result in a larger detergent belt around the protein, unfavourable for crystallization. On the other hand, the use of shorter chain detergents (high CMC) allows more crystal contacts but at the same time increases the risk of protein aggregation due to the risk of exposing hydrophobic segments. Hence, it is crucial to find an optimal detergent for the solubilization, stabilization, and crystallization of IMPs, this is usually performed by screening many types of detergents [36].

A promising approach in facilitating a crystallization event of challenging proteins is the use of high specificity binding partners to form a complex with the target protein. The resulting complex is then used for crystallization trials where the binding partner can offer additional crystal contacts that could increase the likelihood of achieving a higher quality crystal, resulting in higher resolution diffraction. The binding partner can be a fragment antigen binding (Fab), a cameloid single-chain nanobody or lysozyme. In addition, binding partners can be used for stabilizing and capturing the target protein in a specific conformation for crystallization. Some binding partners, such as lysozyme, can even be used for the phasing thereby facilitating structure determination of novel targets [37–39].

It is well established that mobile regions in proteins could prevent the formation of crystals. One way of eliminating these parts is by performing limited proteolysis. This is an easy applicable method where the target protein is incubated with proteases such as trypsin, chymotrypsin or endoproteinase, followed by an analysis through SDS-PAGE. The protease is only able to digest flexible and unfolded parts of the protein that are exposed, generating a protein core that is resistant to further proteolysis, and better suited for crystallization. Limited proteolysis can be combined with Mass Spectrometry (MS) to identify the core sequence, making the generation of a new recombinant protein construct possible [40 41].

The use of commercially available additive screens is an easy and fast way of improving the crystallization quality. Additive screen libraries consist of small molecules such as salts,

amino acids, dissociating agents, linkers, polyamines, osmolytes and reducing agents. These compounds can improve protein crystals by affecting protein-protein and protein-solvent interactions [42].

Another commonly used method to optimize crystal diffraction is crystal dehydration. During dehydration experiments the solvent content in the unit cell is decreased in an attempt to shrink the crystals and hence generate improved crystal packing that could lead to improved diffraction. Dehydration can be performed manually by increasing the reservoir precipitant concentration in a vapour diffusion experiment, pushing the equilibrium towards less solvent in the crystal. It can also be performed in an automated manner at a synchrotron site where an unfrozen crystal is mounted on the Humidity Control Device (HC1b) followed by a controlled decrease of relative humidity around the crystal while the diffraction properties are tested simultaneously [43].

3 From data collection to model building

A tremendous development has taken place in the field of macromolecular crystallography since the first protein structure was solved in 1958. Much of this development concerns computing technology where stronger and faster computers have facilitated improved data collection and processing. The same applies to model building and refinement processes where integrated programs can perform an automated task from processing to model building in basically one click of a mouse. Furthermore, a great development has also occurred around X-rays technologies and most important at the synchrotron facilities. Datasets that used to take several days to collect can now be collected within some minutes using very intense and focused X-rays. These advancements together with improvements in recombinant protein technologies have resulted in an increased number of solved protein structures of targets traditionally considered being challenging, such as IMPs [44 45].

3.1 Why do we need X-rays?

Objects smaller than the wavelength of visible light are impossible to study under a light microscope using traditional scattering. In other words, wavelength of 400-700 nm cannot be directly used to produce an image in the dimension of 1.5 Å which is necessary for accurate determination of protein structures. Therefore, a light source with a wavelength comparable to the smallest features of the object of interest has to be used. In terms of proteins, X-rays are used since they have a wavelength that falls into the range of carbon bonds (1.5 Å). X-rays used in crystallography are highly intense and focused beams, normally produced at synchrotron facilities [34].

3.2 X-ray sources- synchrotrons

In the early years of X-ray crystallography radiation was generated from diverse sources e.g. sealed-tubes and rotating anode instruments. These X-ray sources were relatively limited in both power and precision. Furthermore, the first generation synchrotrons were mainly dedicated for the research field of particle physics, structural biologists could only benefit from the side effect production of X-rays [46]. The second-generation synchrotrons were built

particularly for structural biologists, bending magnets were used to enhance the X-ray quality. However, not until the construction of the third generation synchrotrons could the era of high-throughput crystallography truly advance. Third generation synchrotrons contain a booster ring and a storage ring surrounded by strong magnets and insertion-devises (Figure 11). The production of X-rays in the third generation synchrotrons starts by heating up a cathode until electrons are emitted and transmitted through the injection system into the booster ring. The electrons are then accelerated close to the speed of light before they are sent to the storage ring. Magnets inserted in the ring steer the electrons around and X-rays are emitted every time the electrons are forced to bend. Furthermore, insertion devices that force the electrons to wiggle result in even stronger X-ray emissions. The X-rays are directed from the storage ring to the experimental beam lines where they can be tuned and used for macromolecular crystallography [47, 48].

The next generation synchrotrons could be X-ray free electron laser (XFEL) offering even more advanced applications. This method is based on intense and ultra short X-ray pulses in the range of femtoseconds [49]. The advantage here is the use of micro-crystals, meaning that there is no need of growing large crystals. The micro-crystals are injected into the intense beam through a liquid jet where they are hit by the X-ray pulses. Although the intense X-ray beam evaporates the microcrystals, diffraction information is collected from the intact crystal just before it evaporates [50]. This method is still in the development phase, however the application of XFEL has recently been shown to be very valuable for studies of proteins containing redox metals such as photosystem II [51].

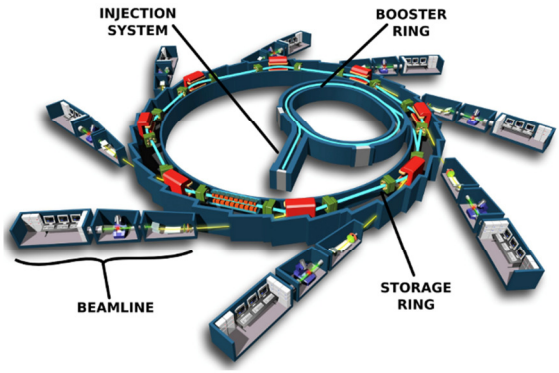

Figure 11. An overview of a synchrotron facility showing the location and the connection between the four major components: injection system, booster ring, storage ring and beam line. Image adapted from the Soleil synchrotron.

3.2.1 Radiation damage

Synchrotron beam intensity is measured in brilliance, taking into account the number of photons that are produced per second per cross section of the beam. The higher brilliance, the more photons reach the crystal and thus the better diffraction. Nonetheless these highly intense X-rays are harmful ionizing radiation. When a protein crystal is exposed to X-rays the electrons in the protein can absorb the photons and gain enough energy to be ejected from the atom. The free electrons then react with atoms to generate free radicals that change the chemical property of the protein in the crystal which often results in poor diffraction due to radiation damage. There are different ways of reducing the radiation damage. One approach is to flash freeze the crystal in liquid nitrogen. In this way, the mobility of the free radicals is slowed down leading to a prolonged lifetime of the crystal in the beam. However, the flash-freezing step can be harmful for IMP crystals since they contain large volumes of solvent. A good cryogenic protector is therefore necessary to protect the crystal from being damaged by ice formation. Moreover, choosing an appropriate data collection strategy is crucial. Here, factors such as beam intensity, beam size, exposure time and detector distance have to be adapted to the crystal size and sensitivity. New synchrotron features such as the helical scan [52], allows for data collection over the entire length of the crystal hence spreading the radiation dose by constantly exposing new parts of the crystal to X-rays. This option is especially important for crystals that are sensitive to radiation damage such as IMPs.

3.3 Bragg's law

When X-rays hit a crystal they diffract in all directions leading to destructive or constructive interference. A constructive interference results in distinct diffraction spots while a destructive interference cancels out hence resulting in no diffraction spots. In 1913 Sir William Lawrence Bragg introduced a way of interpreting diffraction by X-rays. He envisioned the crystal as a set of discrete parallel planes separated by a constant distance (d) where the X-ray beam of known wavelength (λ) hits the crystal planes at an angle of (θ) [53] (Figure 12).

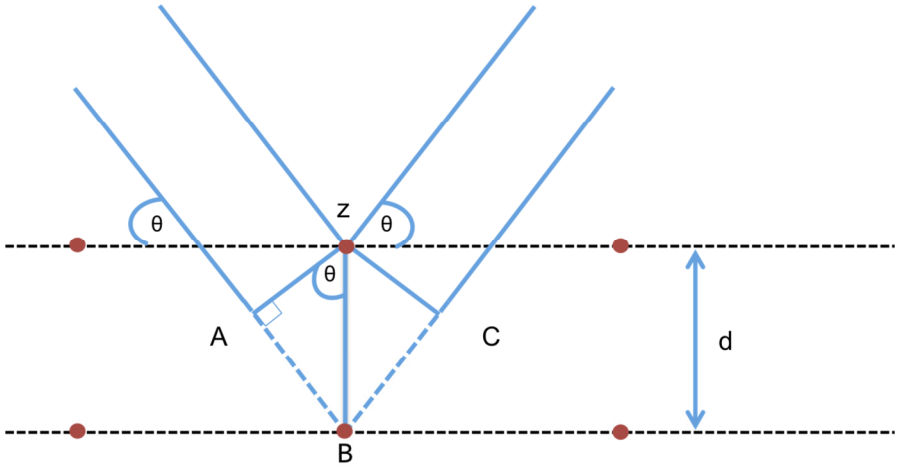

Figure 12. Bragg's law derived by trigonometry gives **nλ = 2 d sinθ**. The blue lines represent two X-ray beams being scattered by two red dots representing atoms. Constructive interference occurs only when the interplanar spacing (d) is equal to an integer multiple (n) of the wavelength.

According to Bragg's law, the distance (d) between the atom planes in a crystal must be equal to an integral (n) multiple of the wavelength for the phases of the two beams to be the same thus leading to constructive interference. The Bragg's equation, **nλ = 2 d sinθ**, shows a relationship between the interplanar distances (d) and the angle θ. The smaller (d), the larger angle of diffraction θ. Therefore, well-packed crystals with tight interplanar distances display higher resolution diffraction [54].

3.4 Data collection

At the synchrotron site, the flash frozen crystal is mounted on a goniometer usually by an automated sample changer. While on the goniometer the crystal is under a constant cryogenic stream of nitrogen gas keeping a constant temperature of around 100 K. Generally, the crystal is first screened by collecting diffraction images from different angles. The diffraction spots are then assigned computationally to determine the space group and quality of diffraction. If the crystal is of sufficient quality, a strategy for data collection is proposed based on those results. During data collection the crystal is rotating on the goniometer allowing X-rays to pass through the crystal in different orientations. A full dataset is collected based on the space group information; the higher the symmetry space group the smaller data collection range is needed for a complete data set. However, for anomalous diffraction measurements more data is collected of each crystal when high redundancy data is preferred [55].

3.5 The phase problem

X-ray diffraction data contain information about the orientation *hkl* and intensities I_{hkl} of each reflection. The electron density p(xyz) can be described by the following equation where the measured intensities I_{hkl} are proportional to the structure factor amplitudes, F_{hkl}:

$$p(xyz) = \frac{1}{V}\sum_{h}\sum_{k}\sum_{l}|F_{hkl}|exp[-2\pi \cdot i(hx + ky + lz) + i\varphi_{hkl}]$$

This equation has to be solved in order to achieve an electron density map. However, there is a problem. The incident X-rays are electromagnetic waves having a known wavelength, amplitude and relative phase, but only the amplitudes can be measured from the diffracted X-rays. Information about the relative phase (φ_{hkl}) is lost and without the phases the electron density equation cannot be solved hence the protein structure cannot be solved. This is the main bottleneck in X-ray crystallography and is widely known as “the phase problem”. Yet, there are some methods by which the phase problem can be solved, they will be discussed below [56,57].

3.5.1 Molecular replacement

Molecular Replacement (MR) is the method of choice for solving the phase problem; it is a straightforward computer-based application where no additional wet lab experiments are needed. However, it does instead require an already known structure (search model) of a homologous protein structurally similar to the unknown target protein structure. As a rule of thumb a sequence identity of at least 25% is recommended for a MR procedure. The lower the identity, the more difficult it is to perform a successful MR. Today, the majority of all structures in the Protein Data Bank (PDB) are phased through MR and as more protein structures are solved, this method will see an even greater application.

In theory, a MR procedure enables phase information to be transferred from the search model to the unknown homologous structure for initial phase calculations. For this to work, the two structures have to be in a similar position in space. This is performed in a two-step procedure, a rotational search followed by a translational search where the aim is to position the search model within the unit cell of the unknown target (Figure 13). Once the search model is positioned in the new unit cell, phases can be calculated and combined with the experimental diffraction intensities to yield an electron density map [58]. However, since most of the structural information is in the phases, there is a risk for model bias in the new structure. Careful analysis and evaluation has to be performed by a so-called omit-map where parts of the structure are omitted in order ensure that they re-appear as a difference density in the new electron density map. Also simulated annealing can be performed to eliminate model bias and to find the global energy minima for the new structure [59].

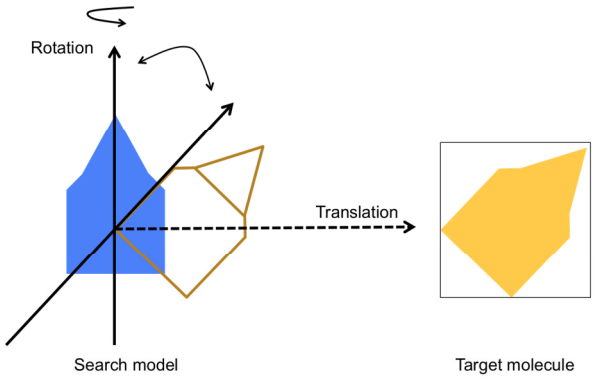

Figure 13. Molecular replacement visualized. The main principle of molecular replacement is the rotation and translation of the search model of the known structure (blue) to align with the target model of unknown structure (yellow). In this way, phases from the search model are applied on the target molecule.

3.5.2 Multiple isomorphous replacement

At the time when the first protein structure was solved, molecular replacement was not a choice. Instead, the phase problem was solved by isomorphous replacement. It was a technique commonly used for small organic molecules but Max Perutz realized that it could also be applicable for large macromolecules. In contrast to MR, multiple isomorphous replacement (MIR) requires additional wet lab experiments to be performed. The underlying idea behind this technique is to incorporate heavy atoms in the crystals. This can be done either by soaking the crystals or by co-crystallization where the heavy atoms bind to the protein by interacting with specific amino acid residues. An important requirement here is to have derivative crystals with heavy atoms as well as a native crystal without any heavy atoms, all crystals must be isomorphous since the addition of heavy atoms must not change the cell-dimensions or the crystal packing. Preferably, a minimum of two different derivative crystals is needed to solve the phases without any ambiguity. Data sets are collected from the native and derivative crystals and since heavy atoms contain more electrons than ordinary protein atoms they will result in a distinct diffraction of higher intensity. From the distinct diffraction it is possible to measure the amplitude, the position of heavy atoms in the unit cell and the phase angle of the wave interfering with the heavy atoms [56]. The accuracy of the predicted phase is validated by different statistical means. The agreement between calculated and observed heavy atom contribution between derivative and native structures is described by the Cullis R factor. Furthermore, the level of phase contribution by each derivative is described by the phasing power and the precision of the estimated phases is described by the figure of merit [55,60].

3.5.3 Anomalous dispersion

Anomalous dispersion is an additional strategy for solving the phase problem where heavy atoms are used without the need for isomorphous crystals since all diffraction data can be collected from a single crystal. The anomalous scattering is achieved by choosing an X-ray wavelength that is close to the absorption edge of the heavy atom. The electrons of the heavy atom will thus absorb and become excited by the X-rays. The excited electrons enter a higher energy shell and when they return to the lower energy shell, radiation is emitted causing a retardation of the diffraction wave. In a Multiple anomalous dispersion (MAD) experiment three data sets are normally collected; one data set at the peak absorption edge of the heavy atom, a second data at the inflections point of absorption and a third dataset at a remote

wavelength. The phases are then calculated based on the differences between structure factors in the data sets [55, 57]. Most heavy atoms used for isomorphous replacement also offer a source of anomalous signal. Combining the two techniques gives SIRAS, Single isomorphous replacement with anomalous scattering, which is an alternative approach to solving the phase problem.

3.6 Refinement

Once the phase problem is solved and an initial electron density map is calculated, model building starts. The first model is usually far from being completed and in order to improve the model quality several iterative cycles of manual model building and refinements are performed, until the data converge. The aim is to build a protein structure that fits into the calculated electron density map as accurately as possible. A successful refinement procedure yields improved phases, giving a better electron density map where a more accurate model can be built and prepared for a new round of refinement (Figure 14). There are several tools available for evaluation of model building quality. One example is by using the Ramachandran plot to examine torsion angles of all amino acids in the model, keeping them in allowed regions is an indicator of a good model.

When building the protein model, many parameters need to be taken into consideration, for example, stereochemistry such as bond length and bond angles of amino acids as well as chemical bond characteristics. At a typical membrane protein resolution of 3 Å, the ratio between collected data and model parameters is not equal; model parameters outnumber the parameters of the collected data. For this reason, restrains and constraints are used in the refinement to improve this ratio. In a restrained refinement certain freedom is given to the geometrical parameters. Restrained refinement therefore results in an elevated number of observations. In contrast, constrained refinement includes strict and exact values for geometrical parameters from which they may not deviate. Constrained refinement therefore results in a decreased number of parameters.

The quality of the refinement is measured by the R-factor (R_{work}):

$$R_{work} = \frac{\sum_{all\ reflections} |Fo - Fc|}{\sum_{all\ reflections} Fo}$$

where Fo are the observed structure factors and Fc are the calculated structure factors. The

R_{work} value is a measure on how well the calculated values agree with the observed values. As a rule of thumb, the R_{work} should be about 10 times the resolution and is usually given in percentage [54]. There is always a risk of over fitting a model into noise, resulting in a false impression of having built an accurate model. For this reason 5-10% of the data are not included in the refinement process, instead they are used for the calculation of an R_{free} value, which is a validation value of the refinement. Throughout the refinement, if the model is correctly built, both R-values should decrease. However, if a model is built into noise, the R_{free} value will not decrease further [60].

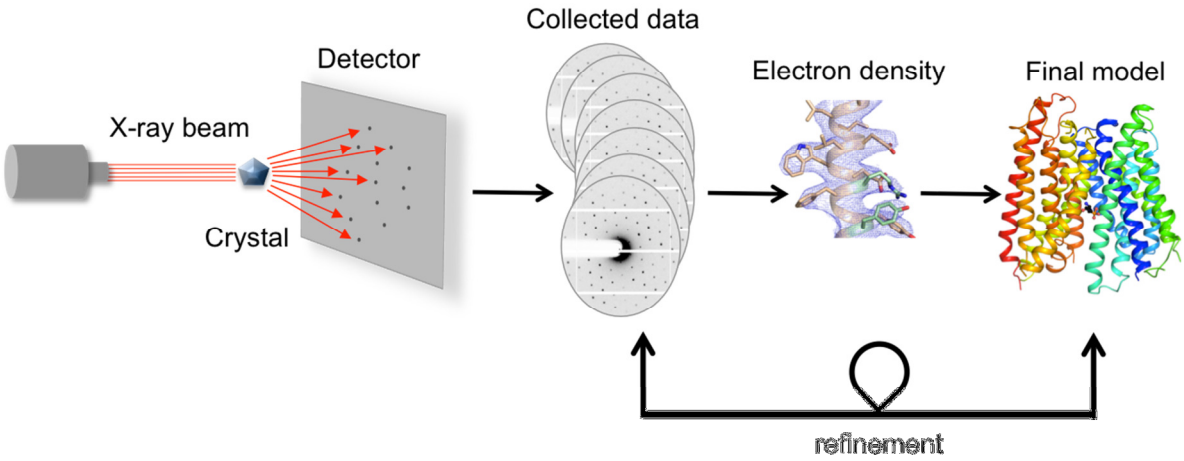

Figure 14. An overview of data collection and model building. X-rays are shot on a protein crystal rotating on a goniometer. The collected data are processed (indexing and scaling) and phases determined before a model can be built. The model building procedure consists of several iterative cycles followed by refinement.

4 Solute transport across membranes

Living cells require a constant flow of molecules across the cell membrane. Nutrients and building blocks are taken up while metabolic by-products and xenobiotics are released. Although some small molecules are able to diffuse freely through the cell membrane, larger and charged molecules are often dependent on special assistance. Therefore, a broad range of transporters and channels have emerged through evolution. These IMP systems enable specific substrates to pass through the membrane, either by facilitated diffusion or by active transport. In this section, channels and transporters will be introduced, followed by an additional focus on transporters.

4.1 Channels

In the early 1950s, two scientists, Alan Hodgkin and Andrew Huxley studied functional aspects of ion channels. They developed an electrophysiology method, called voltage clamp, which allowed them to detect and measure action potentials through the giant axon of veined squid *(Loligo forbesii).* Eventually, these discoveries earned them a Nobel Prize in Physiology or Medicine in 1963 [62,63]. Today, we know that channels are IMPs specialized in facilitating the downhill movement of substrates, an energetically favourable process (Figure 15). The opening and closing of a channel is typically controlled (gated) by either ligands or by variations in the membrane potential. Consequently, channels are grouped and named according to how they are gated.

Ligand-gated channels are typically composed of a ligand-binding site located at the extracellular part of the protein. Ligand binding triggers a conformational change that subsequently leads to the opening of the channel pore. For example, channels located at the neural postsynaptic sites, such as the $\alpha_1\beta_2\gamma_2\gamma$-aminobutyric acid type-A ($GABA_A$) activated channel, belong to this group. A release of $GABA_A$ neurotransmitter activates the channel and a flow of chloride ions is initiated in order to inhibit an action potential from firing [64].

Voltage-gated ion channels hold an intracellular voltage-sensing domain allowing them to open and close upon changes in the membrane potential. This type of channel is essential for the excitation and propagation of the action potential in nerve cell signaling. For example, the voltage-gated sodium channel belongs to this group and is responsible for the rising phase of the action potential [65].

4.2 Transporters

In contrast to channels, transporters can often promote uphill transport of substrates. This type of transport is energetically unfavourable and therefore in need of an energy source, such as ATP or an electrochemical gradient. Transporters typically bind a single substrate followed by a series of conformational changes before the substrate can be transported; consequently this type of transport is slower in comparison to transport through channels. Based on the type of energy source consumed, transporters are grouped into two main groups: primary active transporters and secondary active transporters [66].

4.2.1 Primary active transporters

Primary active transport systems are often transmembrane ATPases, an example is the ABC-transporter which is a primary active transporter belonging to a large protein family with representatives in all kingdoms of life. They utilize the energy from ATP hydrolysis to catalyse the uphill transport of various substrates. Furthermore, they share a characteristic structural architecture of two transmembrane domains (TMDs) and two cytoplasmic nucleotide-binding domains (NBDs) (Figure 15). The alpha-helical TMDs form the pore section and undergo conformational changes during substrate transport. The amino acid sequence of this domain is variable, hence enabling broad substrate specificity. The ATP hydrolysing NBDs, on the other hand, are highly conserved. ABC transporters can be further divided into two subtypes based on the direction of transport: importers and exporters. ABC importers are exclusively present in prokaryotes, and depend on a binding protein (BP) for substrate binding and delivery to the pore. ABC exporters, on the other hand, are BP-independent extrusion systems that are found in both eukaryotes and prokaryotes [67,68].

4.2.2 Secondary active transporters

Secondary active transporters utilize the stored electrochemical gradient over the cell membrane to couple energetically un-favoured transport. They are divided into three groups on the basis of how the energy coupling is accomplished: symporters, antiporters and uniporters (Figure 15) [69].

Symporters co-transport a minimum of two compounds, in the same direction. Typically, ions such as H^+ and Na^+ move down their concentration gradient, allowing a substrate to move up its concentration gradient. The transporter name usually reflects the type of ion being used, as in proton- or sodium- dependent symporter. Proton dependent oligopeptide transporters (POTs) belong to this class of symporters were they use the proton gradient for the uphill transport of di- and tripeptides [69 70].

Antiporters follow the same principle, however the transport of substrates is in the opposite direction. A well-characterized representative of this class is the Na^+/H^+ exchanger (NhaA). This transporter is responsible for the regulation of intracellular pH by exchanging one external Na^+ for an internal H^+ [71].

Uniporters, also known as facilitators, were added to the class of secondary transporters due to structural and functional similarities. They do however differ from the previously described transporters regarding the lack of coupled transport. Uniporters facilitate the downhill transport of substrates thus no energy is needed. The human glucose transporter GLUT1 is a uniporter responsible for the vital diffusion of glucose to the brain and other inner organs [72].

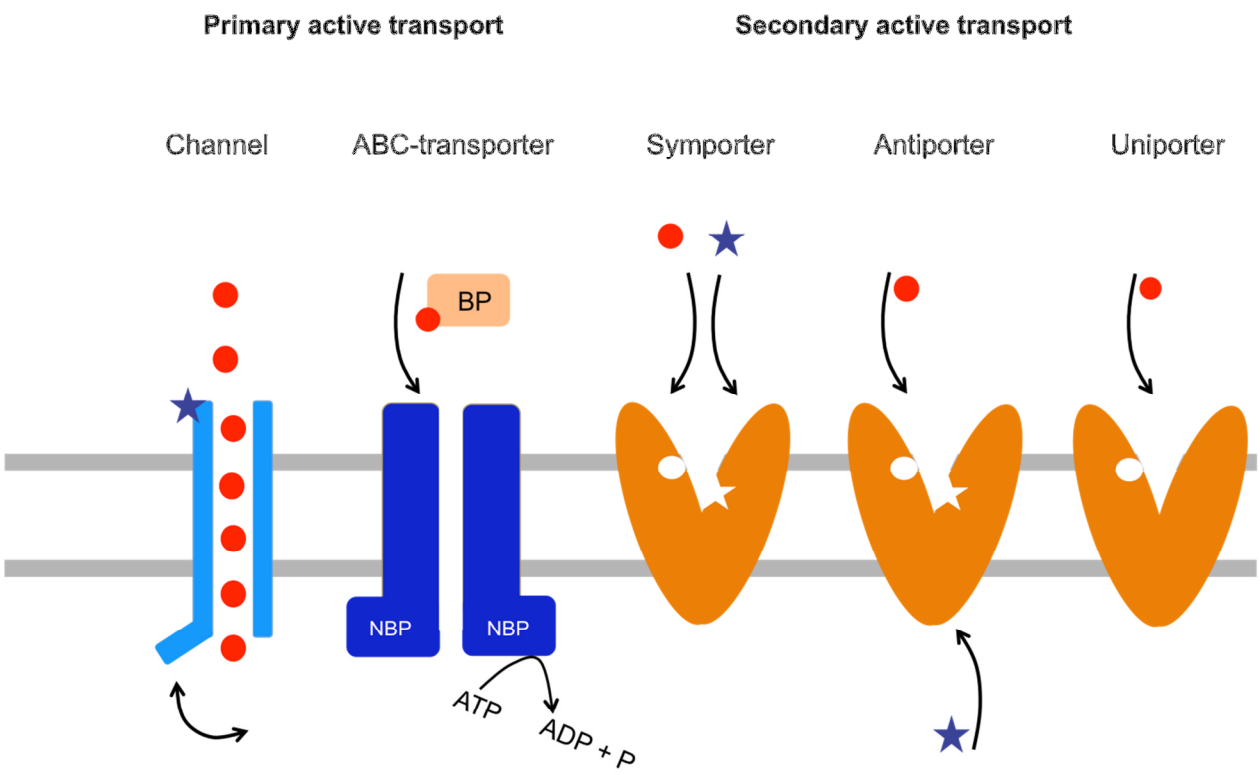

Figure 15. Various ways of substrate transport across the cell membrane.

4.3 Is it a Channel or a transporter?

Sometimes the boundaries between channels and transporters are unclear, especially regarding fast turnover transporters and slow turnover channels. Nevertheless, there are some key differences that may be used as hallmarks for the separation of the two. First, channels interact weakly with their substrates, allowing an unhindered flow down a concentration gradient. Transporters, on the other hand, bind to one substrate at a time and undergo specific conformational changes to perform the uphill substrate transport while consuming energy. Second, transporters alternate between several conformational states during a full transport cycle while exposing the binding site only to one side of the membrane at a time. Using these definitions, some proteins were recently re-grouped. One example is the former chloride-conducting ion channels (CLC) which nowadays is referred to as a H^+/Cl^- transporter [73,74]

4.4 Major Facilitator Superfamily transporters

The major facilitator superfamily (MFS) is the largest group of secondary active transporters, containing over 100 different families [75]. They are found in all kingdoms of life where they are specialized in the transport of various substrates including sugars, peptides, ions, vitamins, and fatty acids. The first structure of a MFS transporter, the oxalate:formate antiporter (OxlT), was solved in 2001 at a resolution of 6.5 Å [76,77]. Since then, additional structures have been solved, mainly from prokaryotic sources such as the glycerol-3-phosphate:phosphate antiporter (GlpT) [78], the D-xylose:H^+ symporter (XylE) [79], the fucose symporter (FucP) [80], the multidrug transporter EmrD [81], the peptide transporters (PepT) [82], and the lactose transporter (LacY) [83]. In recent years, the number of MFS transporters from higher eukaryotes has steadily augmented as a result of new technologies in combination with innovative strategies. The structure of human GLUT1 was solved at a resolution of 3.2 Å [72] and the structure of the fungal phosphate transporter at a resolution of 2.9 Å [84]. Taken together, these structures have revealed a common architecture of MFS transporters consisting of 12 transmembrane helices organized in two 6-helix bundles enclosing the substrate-binding site in the middle. Furthermore, these structures have also contributed to a deeper understanding of the various conformational states in the transport cycle of MFS transporters [85].

4.5 The alternate access mechanism – an overview

A general theory describing the transport-cycle of secondary active transporters was postulated already 50 years ago by Oleg Jardetzky [86]. This theoretical paper states that secondary active transporters facilitate cargo translocation by shifting conformations without opening a continuous pore across the membrane. The transporter must therefore isomerize between at least two distinct conformations, an inward open and an outward open state. This type of transport mechanism is nowadays widely recognized as the a*lternate access mechanism.* So far, several distinct MFS structures originating from different organisms have been solved providing significant evidence for the alternate access mechanism. Apart from structural evidence, the alternate access mechanism has also been studied by biochemical and biophysical assays as well as by computational models [87] [88] [89] [90]

The molecular foundation for the alternate access mechanism is based on structural repeats found in all MFS transporters. The N- and C-terminal domains are related to each other by a pseudo-twofold symmetry axis perpendicular to the membrane plane. This type of symmetry

relationship is frequently encountered among transporters, suggesting a functional importance that might have arisen due to gene duplication. The two sub-domains might be further divided into the smallest primordial repeating unit from which the gene duplication events may have started. Here, two main hypotheses exist. The MFS transporters might have evolved via a gene duplication of a primordial 6-helical domain that in turn evolved either through gene triplication of a two-helix bundles or a gene duplication of a three-helix bundle [75 91].

The structural repeats are stabilized by hydrogen bonds, some of these hydrogen bonds are termed *gates* since they alternate between the state of being broken and re-formed during the transport cycle. There are two groups of gates: intracellular gates and extracellular gates, also referred to as 'outward-facing' and 'inward-facing' gates [92,93]. Altogether, a symporter transport cycle might start in the outward open state, stabilized by the cytoplasmic gates. At this state the binding site is exposed to the extracellular environment and hence accessible for substrates and ions. Once the binding site is fully occupied, a conformational change leads to an occluded state. The occluded state is an important transition that prevents the transporter from forming a leaky channel when switching from an outward open to an inward open conformation. This is the only conformational state where both the intracellular and the extracellular gates are closed. A transformation to the inward open conformation enables substrate and ion release into the cytoplasm thus completing the transport process. Now, the empty transporter cycles back to the outward open state for a new round of transport (Figure 16).

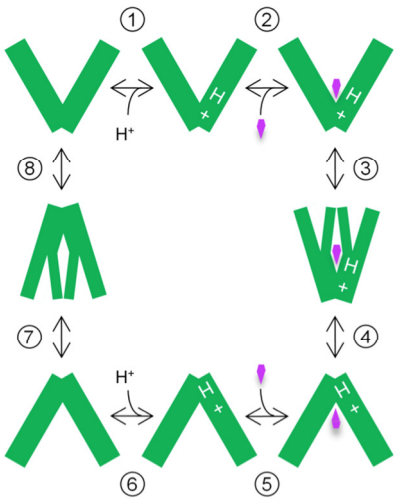

Figure 16. A schematic model of the alternate access transport mechanism of symporters. (1) The transporter is protonated followed by substrate binding (2) leading to an occluded state (3). The transporter then undergoes a conformational change to the inward open state (4), the substrate is released (5) and the transporter is de-protonated (6). The transporter then cycles back to the outward open state, through an additional occluded state, and a new cycle can begin (7-8).

4.6 Proton dependent oligopeptide transporters

E. coli have three systems for peptide uptake: Dipeptide-permease Dpp, Oligopeptide-permease Opp, and Tripeptide-permease Tpp. The names indicate the backbone length of the transported peptides, nonetheless they do show an overlapping selectivity. The Dpp and Opp systems are primary active transporters that are not conserved in humans while the Tpp system belongs to the secondary active transporter family which is conserved in humans [94]. POTs belongs to the Tpp system and are specialized in the uptake of short-chain peptides, an important and energy- providing strategy for the cell to acquire amino acids for downstream cell growth pathways. Furthermore, POTs are conserved in all kingdoms of life, from bacteria to man.

PepT1 and PepT2 are the two human peptide transporters homologous to the prokaryotic POTs. PepT1 is localized in the brush border membranes of the small intestinal epithelium tissue where it is responsible for nutrient uptake [95] (Figure 17 A). PepT2 is expressed in a variety of organs including the kidneys where it is responsible for the reabsorption of peptides hence influencing the *in vivo* half-life time of absorbed peptides [96]. In 1970, two American researchers, John Quay and Laura Foster, showed for the first time that PepT1 and PepT2 are responsible for the transport of short peptides as well as antibiotics [97]. Today we know that these peptide transporter systems recognize and transport a broad variety of di- and tri-peptides including various compounds that resemble small peptides such as β-lactam antibiotics [98]. They are therefore relevant for understanding absorption of orally administrated drugs and for designing drugs with improved absorption and distribution properties. Nevertheless, although PepT1 and PepT2 have been meticulously studied regarding substrate specificity and transport kinetics, there is still no structural information available (Figure 17 B). The lack of structural information is mainly due to the complex and hydrophobic nature of human IMPs leading to complications from protein expression to crystallization. To circumvent this hurdle, prokaryotic homologous transporters are studied since they offer a less complex system providing a good starting point for better understanding PepT1 and PepT2.

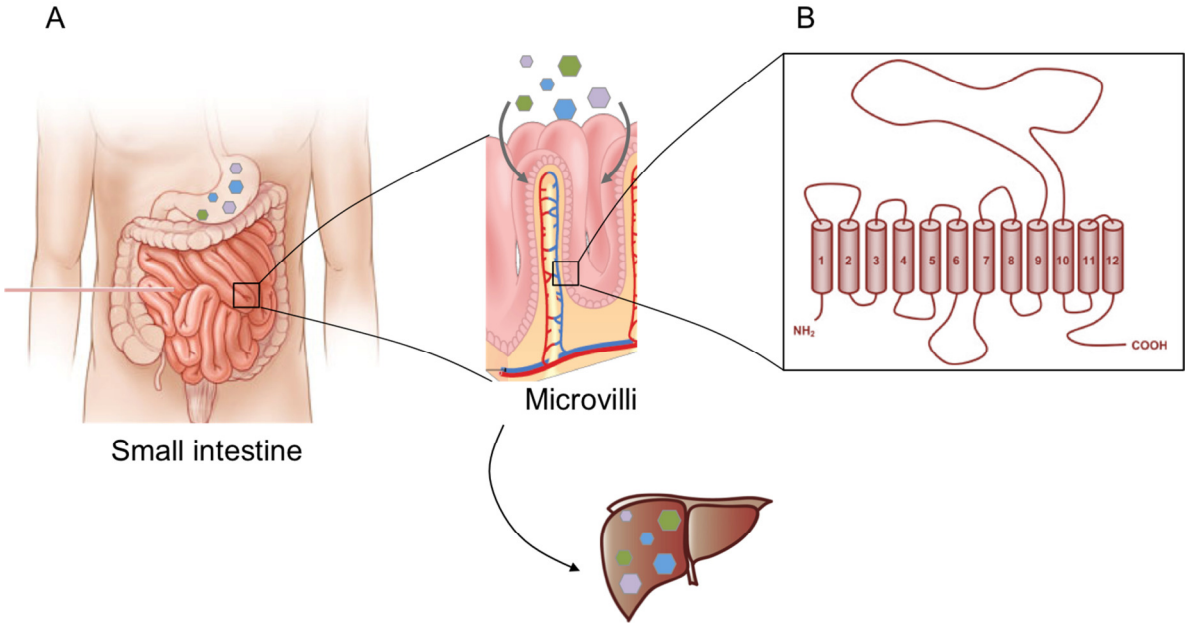

Figure 17. An illustration showing the location of PepT1 in the human body. A) Once ingesting a protein-rich meal, secreted enzymes in the stomach start to digest the protein until a mixture of amino acids and small peptides remain. These short peptides are then transported across the guts and delivered through PepT1 to various organs where they are needed. B) Overall topology model of PepT1/PepT2 showing 12 TM helices and a long segment between TM helix 9 and 10. Figure modified after [99].

4.6.1 The overall architecture of POTs

To date, structures of five different prokaryotic POTs have been determined by X-ray crystallography (Table 1). Those include: $PepT_{So}$ [100] from *Shewanella oneidensis*, $PepT_{St}$ [101], from *Streptococcus thermophiles*, GkPOT [104] from *Geobacillus kaustophilus,* YbgH from *E. coli* $PepT_{So2}$ [82] from *Shewanella oneidensis*. These transporters consist of approximately 500-600 amino acids sharing an overall low sequence identity (around 20%) but high structural similarity. They all exhibit the canonical MFS fold consisting of 12 transmembrane helices having both amino and carboxyl termini facing the cytosol (Figure 18). The helices are organized into two domains, the N-terminal domain (TM helix 1-6) and the C-terminal domain (TM helix 7-12), generating an overall shape of "V/Λ" where the highly conserved substrate-binding site is located in the centre. Furthermore, the two domains are linked together by two additional helices, named HA and HB, sharing low sequence conservation and are absent in PepT1 and PepT2. The function of HA and HB is currently unknown, however there are various speculations claiming a role in stability, oligomerization or regulation of transport [105].

Table 1. Known peptide transporter structures.

POT	Source	Conformation	Bound ligand	Resolution (Å)	PDB code
$PepT_{So}$	*Shewanella oneidensis*	Partially occluded	No	3.62	2XUT
$PepT_{St}$	*Streptococcus thermophilus*	Inward open	Yes /No	2.35 – 3.30	4APS, 4D2B, 4D2D, 4D2D
$PepT_{So2}$	*Shewanella oneidensis*	Inward open	Yes	3.15 – 3.91	4LEP, 4TPG, 4TPH, 4TPJ
$PepT_{GK}$	*Geobacillus kaustophilus*	Inward open	Yes /No	1.90 – 2.40	4IKV, 4IKW, 4IKX, 4IKY, 4IKZ
YbgH	*Escherichia coli*	Inward open	No	3.40	4Q65

Although POTs differ in overall sequence identity across organisms they still share three highly conserved signature motifs [106]. Two of the motifs (ExFxxYGM and FYxxINxG) are specific for the PepT family members while the A-motif (GGxxxDxxxG) is specific for the MFS family. All three motifs are located within the first five helices and are predicted to have different functional roles [107]. A recent paper published by Zhao *et al.* [103] propose that the POT specific motifs are involved in substrate recognition in the binding site, while the MFS specific motif functions as a conformational switch, stabilizing certain conformations in the transport cycle. Furthermore, a mutation in any of these signature motifs results in reduced or lost transport activity.

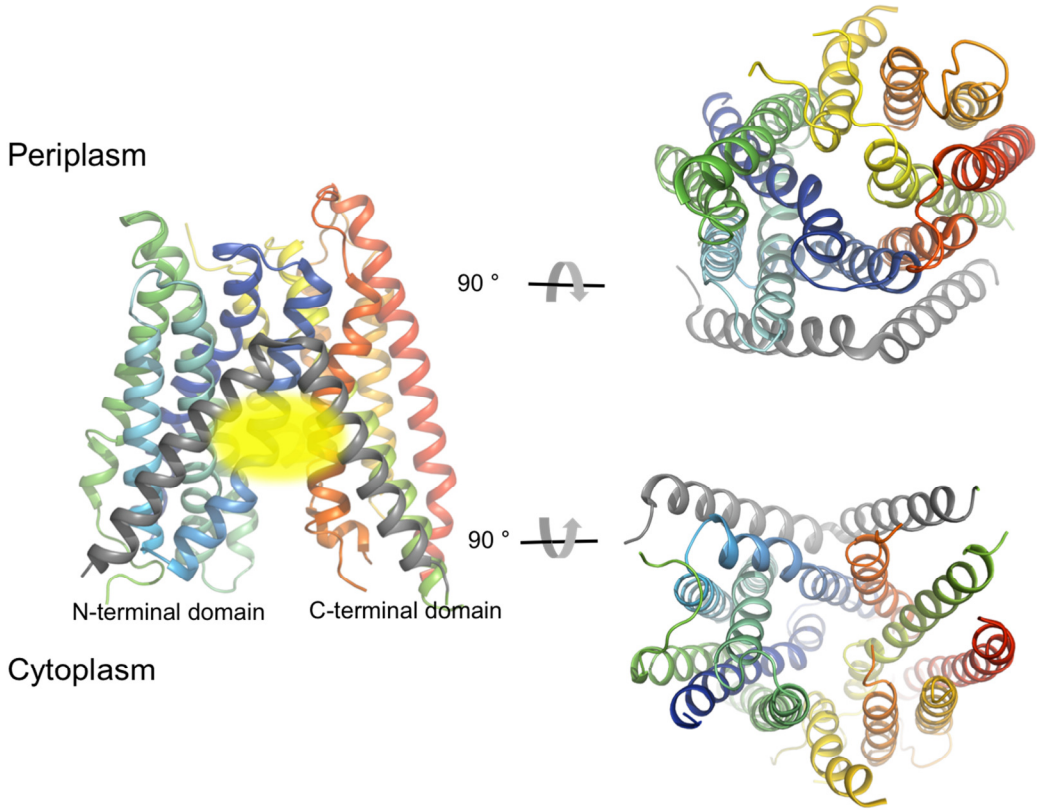

Figure 18. Overall structure of $PepT_{So2}$ visualized from different orientations: side view, periplasmic view and cytoplasmic view. The protein was crystallized in an inward open state where the binding site (yellow) is accessible from the cytoplasm. HA and HB helices are coloured grey.

4.6.2 The binding site

The POT substrate-binding site is a cavity found in a central position between the N- and C-domains. It is the most conserved part of the transporter, sharing an identity of about 90% between human and bacterial POTs. Studying substrate transport of prokaryotic POTs will therefore potentially provide valuable information also regarding the function and mechanism of PepT1 and PepT2.

While studying the structure of $PepT_{So2}$ in paper II-III, we discovered a dipole-like charge distribution along the binding site by where the substrates were bound. The substrate N-terminus interacted with the negatively charged part of the pocket consisting of E402, N151 and N329, while the substrate C-terminus interacted with the positively charged part of the pocket consisting of R25 and K121 (Figure 19 A). One of the main mysteries of the binding site is the broad substrate specificity. How is it possible for a single site to recognize and transport substrates ranging from di- and tri-peptides to drugs? Some clues can be found when studying the structural architecture of the binding site. A striking feature is the large number of tyrosine residues in the binding site and in its vicinity. It is likely that tyrosines play an important and versatile role in the substrate coordination since they are able to interact with the substrate through various types of interactions such as hydrogen bonds, hydrophobic and electrostatic interactions. Furthermore, tyrosine residues are also able to switch rotamers hence adapting the size of the binding site to the size of the substrate. The plasticity of the binding site is further amplified by the strategic location of R25 and K121, enabling peptides of different sizes to interact through their C-terminus (Figure 19 B).

In 2013 Keisuke Ito *et al.* [108] investigated the substrate specificity of Ptr2p, a POT from *Saccharomyces cerevisiae*, by using a dipeptide library. The results showed a remarkable substrate preference biased towards di-peptides consisting of essential and semi-essential amino acids. In 2014, we investigated the substrate specificity of $PepT_{So2}$ was investigated by a proteoliposome based uptake assay of tripeptides (Paper III). Results from this study showed a preference for small and medium sized hydrophobic residues but a disfavour for glycines [109].

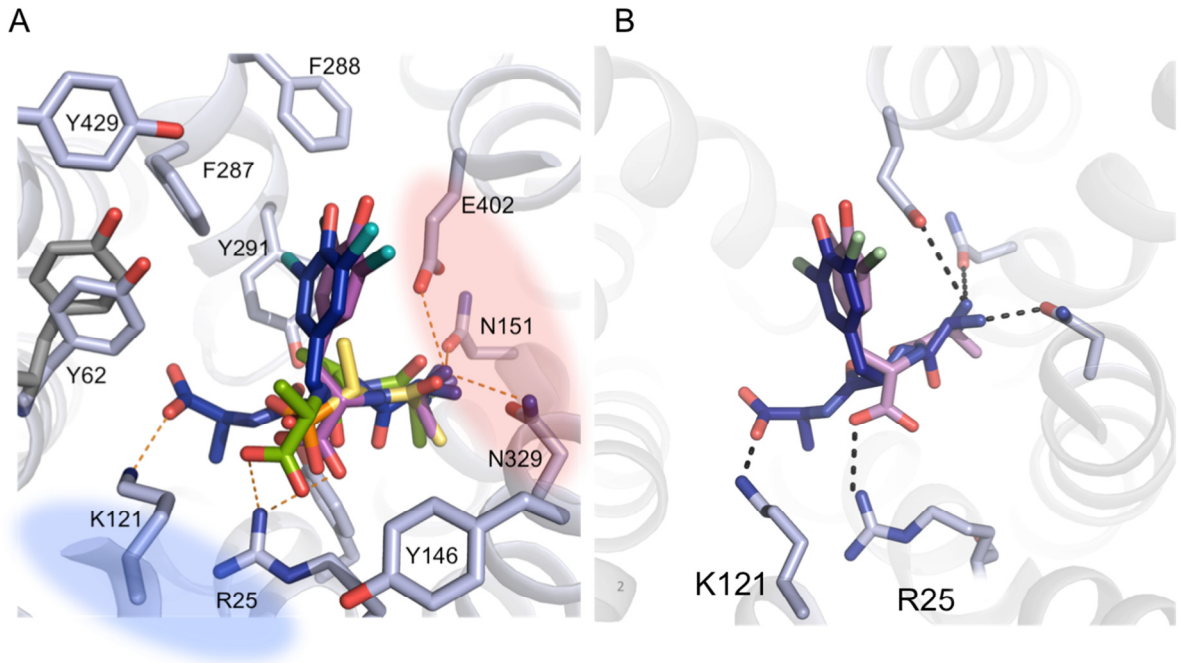

Figure 19. A) A close up view of the $PepT_{So2}$ binding site where three structures with substrates are aligned (AAA, AY and AYA). Conserved side chains responsible for coordinating the substrates are illustrates as sticks and the dipole like charge distribution is illustrated as red and blue patches. B) Differences in the coordination of substrate C-termini by R25 and K121.

4.6.3 Proton coupling

As mentioned earlier, peptide transporters are secondary active symporters utilizing the electrochemical gradient stored over the cell membrane to energize for uphill substrate transport. A prerequisite for such a transport system is a tight membrane not permeable for unspecific ion leakage, enabling ion gradients to be maintained and used in a controlled manner to drive metabolic processes. Different types of ion gradients are present over the cell membrane such as Na^+, Cl^-, and H^+. However, the most commonly used ion gradient by peptide transporters is the proton gradient (H^+) generated by specific membrane embedded proton pumps such as the ATPase and the respiratory chain proteins. Since the proton gradient is composed of both a charge ($\Delta\Psi$) and pH (ΔpH) it is usually referred to as the electrochemical force.

It is generally accepted that for secondary active symporters, the protonation event occurs at titratable residues in the vicinity of the substrate-biding site. Nevertheless, the molecular details are still obscure. The most studied and best-understood MFS transporter, LacY, undergoes a protonation event prior to sugar binding. Furthermore, the substrate:proton ratio

was predicted to be 1:1 [110] and the release of the sugar was preceded by a de-protonation event [111]. Daohua Jiang *et al.* investigated the proton coupling mechanism of the putative drug efflux protein YajR [107]. According to their hypothesis, the outward open transporter binds a substrate prior to a protonation event, in this way preventing protons from leaking through the transporter. Also, an inward open YajR releases the substrate prior to a deprotonation event, thus preventing the substrate from being transported back. In a recently reported study of the peptide transporter YbgH [105], it was concluded that protons bind in response to substrate binding. Furthermore, the protonation site was predicted to be a cluster of residues, located in the vicinity of the substrate-binding site, serving as a sensor for substrate binding. Similarly, in a study of the cation/melibiose symporter, MelB, an analogous conserved cluster was identified as the protonation site [112]. A study published by Parker *et al.* [113] provide some hints on the proton:substrate stoichiometry of $PepT_{St}$. Here, the proton-coupling stoichiometry was studied mainly by a reconstituted proteoliposome system in combination with radioactively labelled substrates or a pH sensitive dye fluorophore. Based on these experiments it was proposed that POTs use different proton stoichiometry depending on what type of substrate is being transported, 3:1 for tri-peptides and 4-5:1 for di-peptides [113]. Taken together, these results point towards significant variations in proton coupling and substrate translocation among and within MFS transporters. The proton coupling by $PepT_{So2}$ is still not well understood. The protein has three highly conserved titratable residues in the binding site: Glu21, Glu24 and Glu409. From the substrate bound structures, in $PepT_{So2}$, we predicted that Glu 402 was involved in binding of the substrate N-terminus, leaving Glu 21 and 24 as the candidates for proton coupling. Interestingly, mutagenesis studies on the corresponding residues in $PepT_{St}$ showed a loss of function emphasizing the importance of these residues [114].

5 Summary of papers

5.1 PAPER I

High-throughput analytical gel filtration screening of integral membrane proteins for structural studies

Paper I describes a rapid and cost-efficient screening procedure that enables fast identification of protein constructs, expression strains and purification conditions for structural studies of IMPs. In this workflow, we screened the expression of prokaryotic IMPs in *E. coli.* However, the overall procedure is also applicable for other target proteins or protein complexes produced in other expression systems.

The screening strategy was designed as a pipeline divided into the following steps: construct design, protein expression screening, protein purification via IMAC and SEC, and finally crystallization. Furthermore, checkpoints with different criteria had to be fulfilled before a target could enter the next step. Selection of successful targets was based on the following criteria: correct DNA sequencing, expression yields, detergent solubilization potential and homogeneity of purified protein, along with crystallization hits in broad screens (Figure 20).

A total of 48 prokaryotic IMPs were selected based on structural and functional novelty and on the existence of human homologs. Target DNA was cloned into pET vectors using the ligation independent cloning (LIC) system creating two constructs of each target, carrying a C-terminal or an N-terminal His-tag. Out of 96 possible constructs 93 were found to have a correct DNA sequence and were thus sent forward to the protein expression-screening step.

Two *E. coli* strains were used for the small scale expression tests, BL21(DE3) and C41(DE3). The C41(DE3) strain belongs to the Walker strains and is optimized for overexpression of membrane proteins [115]. Crude membrane preparations were used for evaluating protein expression levels using Western blot analysis in comparison to a reference protein. A minimum threshold for the expression level was set and targets expressing above this threshold were subjected to small-scale purification.

The small-scale purification procedure included the use of IMAC followed by SEC. Since the detergent choice is important for IMP homogeneity and stability, we performed detergent screening including five different detergents commonly used for membrane protein solubilization (FC12, DDM, LDAO, Cymal-5). Each membrane fraction was thus solubilized in those detergents prior to purification. Target purity and homogeneity was then evaluated by SDS PAGE and SEC elution profiles.

Based on the small-scale purification results, targets were selected for large-scale expression and purification. The large-scale expression and purification strategy was performed in the same manner as the small-scale screening. Targets that eluted as monodisperse peaks from the SEC were concentrated and used for crystallization trials.

A summary of the results from the screening pipeline is shown in Figure 21. Interestingly, Western blot analysis revealed higher protein expression levels for N-terminally His-tagged targets expressed in BL21(DE3) compared to C-terminally His-tagged proteins expressed in C41(DE3) cells. However this trend was reversed in the small-scale purification step. Here, targets expressed in *E. coli* strain C41(DE3) with a C-terminal His-tag eluted more often with a monodisperse SEC peak indicating a homogenous sample compared to targets expressed in BL21(DE3) with a N-terminal His-tag. A possible reason for this observation might be that adding an N-terminal His-tag could interfere with protein recognition and insertion into the membrane. Furthermore, measuring protein expression levels with Western blot analysis only gives a hint on the expression level, it does not give any information about protein quality and stability. A lesson learned from this screening pipeline was the successful combination of a C-terminal His-tag construct expressed in the C41(DE3) strain and solubilized in DDM.

In conclusion, the screening pipeline enabled us to rapidly screen 48 IMPs and to identify candidates for structural studies. Using identical conditions for the small scale and the large-scale procedure greatly accelerated the progress from analytical to preparative results, while ensuring a high success rate for the transition. Furthermore, the easy setup and low-cost permit this method to be used in many labs, where various targets and expression systems can be applied. Using the described procedures we were able to identify successful constructs and expression strains for various IMPs. In addition, we also generated crystals and collected initial diffraction data for six transporters: $PepT_{So2}$, $PepT_{St}$, a MATE transporter, an arsenate transporter, a folate transporter, and a xylose transporter. This underlines the efficiency and potential value of this type of screening procedure, which is of great utility, since large scale expression, purification and crystallization of IMPs is both challenging and labour intensive.

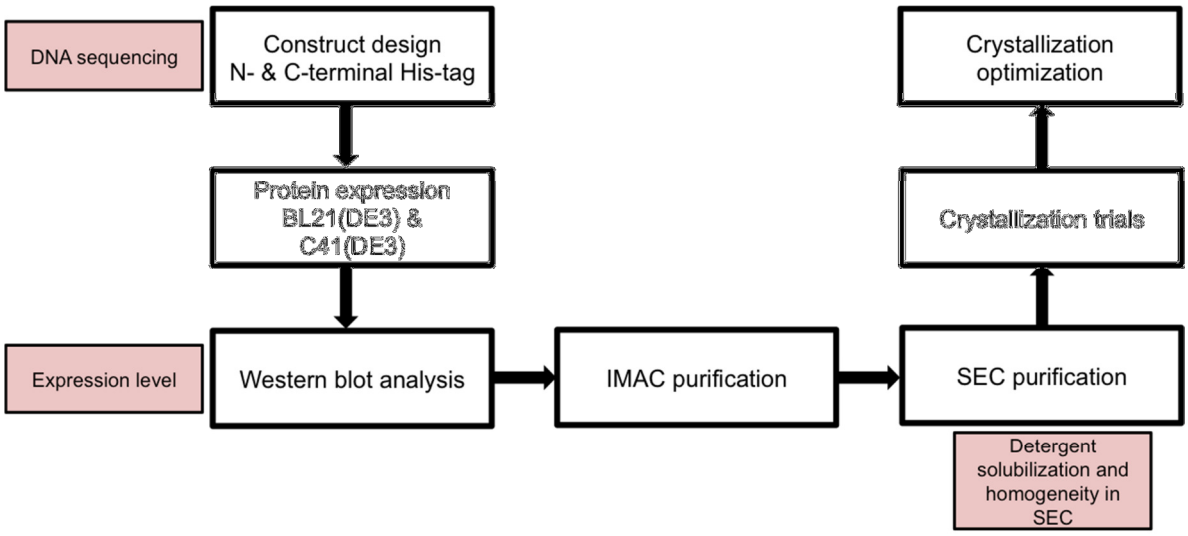

Figure 20. A flow chart of the screening pipeline, describing the pathway from construct design to crystallization optimization. The checkpoints are shown as coloured boxes.

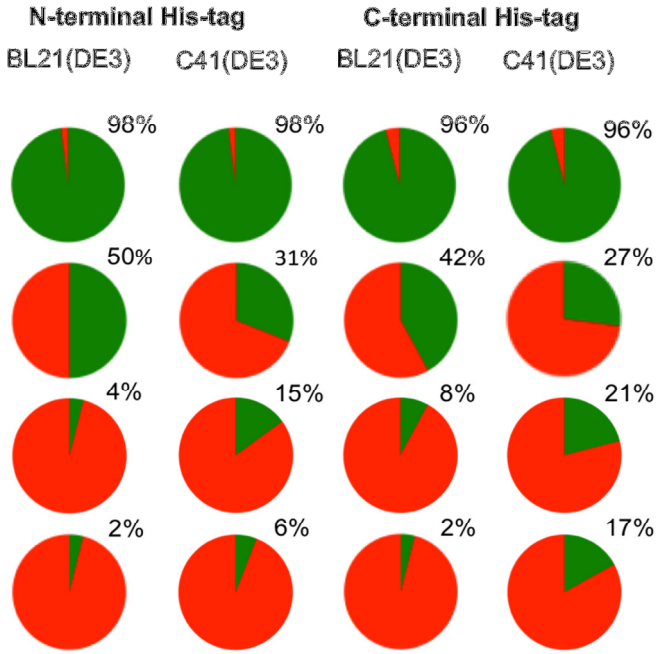

Figure 21. Overview of the success rates in the screening pipeline. The circle diagrams show a percentage of targets passing through each phase (green) relative to the number of targets entering the first phase.

5.2 PAPER II-III

- **Structural insights into substrate recognition in proton dependent oligopeptide transporters**
- **Selectivity mechanism of a bacterial homolog of the human drug-peptide transporters PepT1 and PepT2**

Based on the workflow developed in paper I, we identified a proton dependent peptide transporter from *Shewanella oneidensis* ($PepT_{So2}$) as a promising target for structural and functional studies. We were very interested in the structural basis for the multi-specificity of peptide transporters. In paper II and III, we present the inward open crystal structure of $PepT_{So2}$ in complex with four different substrates: an antibacterial peptidomimetic (alafosfalin), a di-peptide (AY), and two tri-peptides (AAA and AYA). All structures revealed a consistent substrate coordination mode guided by the dipole-like charge distribution in the binding site. The substrate backbones extend laterally across the binding site enabling the substrate N-terminus to interact with a negatively charged area, while the substrate C-terminus interact with a positively charged area (Figure 19 A). This observation confirms the already observed importance of both the N- and C-termini for substrate recognition and coordination. Furthermore, we observed that side-chains of tyrosine residues of substrates AY and AYA were interacting with a hydrophobic pocket found in the centre of the binding site.

We realized early on in the crystallization process that $PepT_{So2}$ generated high quality crystals only when co-crystallized with a substrate and divalent ions such as zinc. Following these findings initial $PepT_{So2}$ crystals were generated and after several rounds of optimization, a condition was found where reproducible $PepT_{So2}$ crystals grew and diffracted to 3 Å. The diffraction was anisotropic meaning that it reached to the highest resolution diffraction only in certain directions. This feature is very common among IMP crystals and is likely to be connected to their crystal-packing type. The dataset was processed in space group $P2_12_12_1$ using anisotropic scaling with ellipsoidal truncation. The asymmetric unit consisted of two molecules held together as an artificial dimer by a zinc ion (Figure 22 A).

Based on our gathered structural knowledge of $PepT_{So2}$, we divided the binding site in three pockets assigned to the interaction sites of the different substrate side chains. We thus aimed to obtain further insights regarding the notable broad substrate specificity of POTs, by mapping out preferred substrates. Thermal shift assays in combination with proteoliposome

uptake assays were performed using a tri-peptide library. Results emerging from this study confirmed that small and medium sized hydrophobic side chains were generally preferred by $PepT_{So2}$. We also observed that glycine residues were strongly unfavourable.

Interestingly, a POT structure (GkPOT) from the prokaryotic organism *Geobacillus kaustophilus,* also co-crystallized with alafosfalin, was published simultaneous to our structure [101]. In contrast to our study, the GkPOT protein carried a point mutation (E310Q) in the binding site. The authors argued that this mutation was necessary in order to capture alafosfalin in the binding site by mimicking the protonated state. Comparing the alafosfalin coordination in $PepT_{So2}$ with the one of GkPOT reveals some intriguing differences. Although the general alignment is in agreement with the lateral dipole charge distribution, the substrate coordination is quite different. The alafosfalin molecule in GkPOT is located closer to the periplasmic side of the binding site compared to $PepT_{So2}$. Furthermore, the difference was more pronounced at the C-terminal phosphate head group than in the N-terminus (Figure 22 B). The observed shift in coordination could simply be explain by different transport modes but could also be due to the point mutation introduced in GkPOT.

An additional study of a POT transporter co-crystallized with substrates was performed by Lyons *et al.* [114]. Here, $PepT_{St}$ from the prokaryotic organism *Streptococcus thermophiles* was co-crystallized with the peptides AF and AAA. While the AF peptide was coordinated in a similar fashion to our di- and tri-peptides i.e. in accord with the dipole moment of the binding site, the AAA coordination eas dramatically different from the common binding mode. The AAA substrate was instead bound in a vertical position where the C-terminus extends towards the periplasm. Thus, although $PepT_{So2}$ and $PepT_{St}$ share a highly conserved binding site, the coordination of AAA is strikingly different (Figure 22 C). The authors speculate that this new binding mode enables $PepT_{St}$ to coordinate a broader range of substrates. Moreover, although the structural resolution of the $PepT_{St}$ was 2.5 Å, the substrate electron density was weak. Hence, the relevance of the proposed AAA binding mode needs further examination.

An exciting feature revealed in Paper II was the tetrameric state of $PepT_{So2}$. While purifying $PepT_{So2}$ and other POT homologs by SEC we noticed an earlier retention time of $PepT_{So2}$, indicating a larger protein assembly. This was further analysed via Blue Native PAGE, cross-linking with glutaraldehyde and negative stain electron microscopy. Results from these methods confirmed a tetrameric organization of $PepT_{So2}$ while the other characterized POTs, $PepT_{So}$, YjdL and YdgR were monomeric, except for $PepT_{St}$ that might be a dimer. Further evidence for a homo-tetrameric $PepT_{So2}$ was achieved through a low-resolution crystal structure revealing the overall assembly of the tetramer (Figure 22 D). This homo-tetrameric

state forms the basis for further investigations where a higher-resolution X-ray or Cryo-EM structure would shed more light on the molecular details of this arrangement. Furthermore, biochemical assays are necessary to unravel whether the subunits in the tetramer function in a cooperative or independent mode. Results gathered in these two papers together give new important insights into substrate recognition, substrate preference and the quaternary state of POTs.

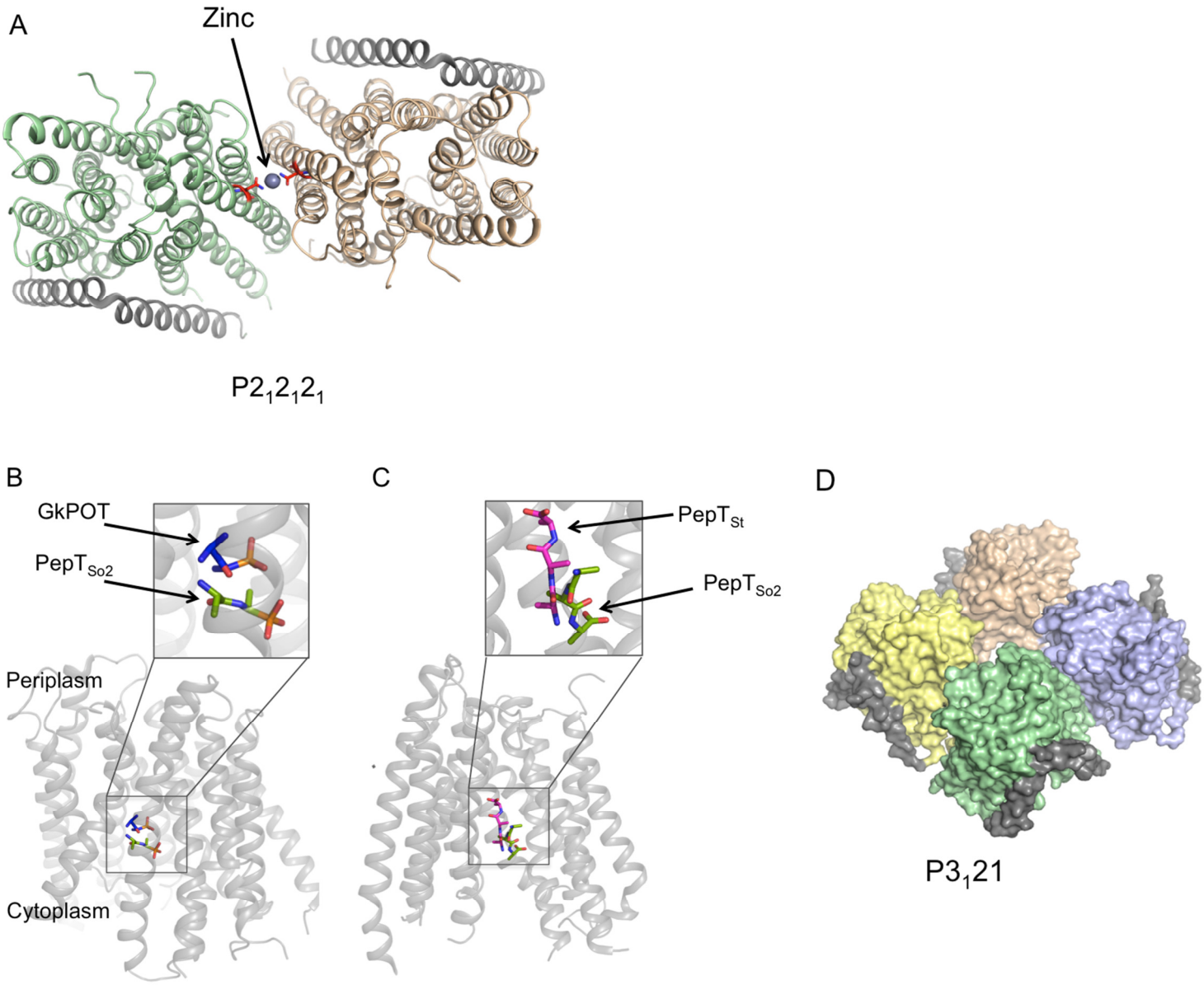

Figure 22. A) The artificial dimer crystal packing obtained from the high resolution $P2_12_12_1$ crystal form. The two molecules are coloured differently and the HA-HB helices are coloured in grey to facilitate the visualization of the symmetry. The dimer interface is stabilized by a zinc ion coordinated by two asparagines, shown as stick s and coloured in red. B) A comparison between alafosfalin position in $PepT_{So2}$ and GkPOT. $PepT_{So2}$ is shown as a grey cartoon and the alafosfaline molecules are annotated and shown as sticks. C) A comparison between AAA coordination by $PepT_{So2}$ and $PepT_{St}$. $PepT_{So2}$ is shown as a gray cartoon and the AAA peptides are annotated and shown as sticks. D) The low-resolution homo-tetramere structure of $PepT_{So2}$ obtained from a $P3_121$. The four units are coloured differently and the HA-HB are coloured in grey to facilitate the visualization of the symmetry. D) The homo-tetramere structure of $PepT_{So2}$ obtained fro a $P3_121$ crystal form at a resolution of 4.6 Å. The four units are coloured differently and the HA-HB helices are coloured in grey to facilitate the visualization of the symmetry.

6 Future Perspectives

Ever since the human peptide transporters PepT1 and PepT2 were identified as drug uptake systems, there have been a large number of scientific studies to unravel their molecular mechanisms. However, human IMPs are not only challenging to produce but also difficult to keep in a stable and functional state leading to extreme difficulties in handling and crystallizing them. The field has therefore turned towards the more stable prokaryotic homologs (POTs). Due to the high conservation they can be used as interesting model systems to study mechanistic details, also relevant for the human proteins. An alternative possibility is the "humanization" of prokaryotic POTs through key mutations, generating a transporter more similar to the human transporters. The mutants could then be used as a platform for understanding of the human transporters.

The field of POTs has in recent years experienced great progress, several high-resolution structures were determined, some of them even bound to substrates. These structures enabled us to, for the first time, visualize molecular details of these transporters, identify key residues in the binding site, and monitor conformational changes. Nevertheless, in spite of the abovementioned novel insights, key questions still remain to be addressed, particularly concerning the proton coupling mechanism, structural insights into various conformations, the kinetics of the transport event, transport regulation, and the quaternary state.

During my PhD thesis, several attempts were made aiming at crystallizing $PepT_{So2}$ in various conformations. Some of the techniques performed include co-crystallization with specific binders such as inhibitors and nanobodies. Furthermore, mutations were inserted at specific gating residues in order to lockup the gates and stabilize a particular conformation. Other approaches such as crystallization in different detergents, in lipid cubic phase, and in lipidic bicelles were also attempted. Although some of these methods yielded initial crystallization hits, more work is required for further optimizations.

To tackle the tertiary structure of $PepT_{So2}$, an in-house nanoparticle technology was developed [116]. This novel approach is based on the lysosomal saposin protein, known for having a lipid modulator activity [117]. They are composed of amphipathic helices enabling the assembly of nano-disc structures where hydrophobic molecules can be incorporated. The tetrameric $PepT_{So2}$ was successful incorporated in the saposin nanodiscs and has proven to be

a promising target for Cryo-EM studies. So far, an initial low-resolution tetrameric structure was obtained and more particles will be collected to improve the resolution.

One of the most applied methods here is the X-ray crystallography. This is also one of the main challenges in the field of IMPs. It is a very tedious and labour-intensive process based on a huge number of crystallization experiments where thousand of conditions and crystals are screened before a well diffracting crystal is identified. The pathway from a purified IMP to a well diffracting and reproducible crystallization condition is thus very long and complicated. Furthermore, the diffraction of IMP crystals is often anisotropic and of poor resolution. This imposes huge difficulties for the model building process that has to be performed by many iterative cycles. It is therefore of great interest to find new and alternative methods that can enable a faster and easier way of solving IMP structures. The lipid cubic phase (LCP) crystallization method is an alternative crystallization method where the IMP is crystallized in a membrane-mimetic matrix. This technique has gained a lot of attention recently due to success in delivering high-resolution structures of various IMPs, including human targets. Another alternative method is the use of X-ray free-electron laser (XFEL) an innovative new opportunity where structural information is collected from micro-crystals. However, the advantage of this method compared to current strategies for data collection need to be proven. The most exciting method for IMP structure determination is probably Cryo-EM. Here, the crystallization step is bypassed since structural information is collected straight from a protein sample without the need of solving any phases. Furthermore, these experiments have the potential to monitor several conformations of the protein in a single experiment. Several IMP structures have recently been determined using Cryo-EM and it will be very interesting to see how generally applicable it is for future routine structure determinations of IMPs.

For many years structural biology of membrane proteins was seen as a very challenging field. However, due to the great dedication from researchers all over the world towards developing more efficient and robust strategies for IMP production, crystallization and structure determination the future of IMP structural biology is bright. In the near future we will likely experience a continued rapid increase in solved membrane protein structures including human IMPs which will enable us to understand the molecular mechanisms of a larger fraction of the biology of the membrane, as well as utilize the structural information to enable more efficient development of novel drugs.

7 Svensk populärvetenskaplig sammanfattning

En levande cell kan liknas vid ett hus där väggarna är cellmembranet och dörrarna är integrala membranproteiner, en sorts proteiner som sitter i cellmembranet och har kontakt med både cellens insida och dess omgivande miljö. Cellmembranet skyddar och avgränsar cellinnehållet från omgivningen medan integrala membranproteiner ansvarar för bland annat kontroll och reglering av molekyltransport över cellmembranet samt kommunikation med omgivningen. Varje molekyl som ska in eller ut ur cellen har tilldelats ett specifik integralt membranprotein, en transportör, genom vilken den tillåts passera.

Denna studie behandlar en sorts integrala membranproteiner som tillhör kategorin peptidtransportörer. De finns i tunntarmens vägg där de ansvarar för absorption av intaget födoprotein och möjliggör på så sätt näringsupptag vid förtäring. Intressant nog är många peptidtransportörer nödvändiga även för upptag av olika typer av läkemedel, till exempel antibiotika. Ökade kunskaper om dessa proteiner och deras egenskaper som läkemedelstransportörer är därför av stor betydelse för framtagning av läkemedel som kan tillföras kroppen via oralt intag. I detta arbete har jag velat studera och förstå hur peptidtransportörerna ser ut, deras tredimensionella proteinstruktur. Utan en proteinstruktur kan man endast spekulera om hur vissa läkemedel tar sig igenom transportören och varför vissa läkemedel inte kan ta sig igenom. Den kunskap som en proteinstruktur tillför läkemedelsutvecklingen kan spara år av arbete och möjliggör riktad design av läkemedel som har rätt egenskaper för att kunna passera igenom transportören på ett effektivt vis.

Jag har under min doktorandtid fokuserat på att förstå funktionen och strukturen på peptidtransportörer som härstammar från olika organismer. Den vanligaste metoden för proteinstrukturbestämning kallas röntgenkristallografi och bygger på att man först genererar kristaller från en lösning av framrenat protein och sedan skjuter röntgenstrålar genom dessa. Röntgenstrålarna sprids när de passerar kristallen och bildar ett mönster som fångas av en detektor. Detta s.k. diffraktionsmönster innehåller information om proteinets atomära struktur. Vägen från idé till ett framrenat protein och vidare till en färdig struktur är emellertid lång och tidskrävande, särskilt vad gäller integrala membranproteiner Våra strukturella studier på peptid transportörer har tillåtit oss att få en inblick i hur dessa proteiner ser ut och hur de fungerar. Detta kan vara första steget mot en mer rationell design av läkemedel som tas upp effektivare av kroppen. I framtiden vill man även designa så kallade ”pro-drugs”, läkemedel som man kopplar till en aminosyra för att göra dem mer peptidlika. På så sätt kan läkemedelsmolekylerna lättare följa med ut i blodcirkulationen.

Acknowledgements

Dear reader, thank you for reading my thesis. You now deserve to read the most popular page.

Working with this project has been a great adventure and I enjoyed (almost) every second of it. The results presented here were generated through a team effort and I was lucky to be part of the best team in the word. I would therefore like to thank you all, with special thanks to:

Pär Nordlund, my supervisor. Thank you for offering me a great PhD environment. You have provided me with all resources I could wish for. I have learned a lot form you, but most of all I have learned to be independent and to aim high.

Christian Löw, my co-supervisor. I don't know how my PhD would have ended up without you! Thank you for sharing your knowledge and for teaching me everything in the lab. I really appreciate the support you give me.

Pelle Always there and always ready to help with a smile on your face. You are one of a kind! Thank you for taking care of everything and for always being ready for a fika. It was an honour working with you.

Esben, thank you for teaching me everything from cray-fishing to crystallography. I enjoyed your company during our looong days in the cave and I admire your calm and humble attitude.

Lionel, you are a true inspiration, thank you for sharing your knowledge about everything. Also, thank your for teaching me crystallography and always having your office door open, both for questions and for laughter.

Jens, what would the lab be without your crazynessss? Thank you for all the guidance, both in the lab and for my future career. Also thank you for all the pep-talks and for comforting me through tuff times with cute Youtube animals.

Michael, Thank you for making the last project more enjoyable and for teaching me the proteoliposome assay. Above all, thank you for all the fun: the running club, the US road trip and the chill in the kitchen. See you in Munich!

My sister PhD students, **Rebecca** and **Rachel**. Thank you for making the office an enjoyable place, for the help whenever needed and all the fun discussions. It was great sharing my PhD adventure with you!

Mikaela, where should I start? It all began during that synchrotron trip and ended up with us sharing beds all over the world. I would not have survived these years without your sense of humour and your warm and kind attitude. You are a true inspiration and my life mentor.

Rozbeh, you were the missing link to kexet and kladdet. Thank you for bringing us together and for creating the best Oreo in the word! I really enjoy your company, what would fika be without your politically incorrect comments? I also really appreciate all your help and support and especially your supply of good quality chocolate!

Daniel, Catrine and **Michael** thank you for a great time and good luck in the future!

The Divne lab: **Christina, Tien-Chye Tan (TC) Rosaria** and **Noor**. Thank you for contributing to the nice atmosphere in the lab. A special thanks to my joyful lab bench neighbours TC and Rosaria for always being ready to help no matter what time or day.

I would also like to thank previous lab members: **Marina** for great supervision during my master's thesis. **Marie, Malin and Denisa** for all the nice discussions always leading to laughter. **Madhan** my master thesis fellow, I will always remember your manly-dance!

A big good luck to the new lab members: **Sara, Anette, Sylvia and Johan**.

The Schüler team: **Herwig**, **Aja, Torun, Tobias, Filipa**. Thank you for contributing to the great atmosphere here.

The PSF team: **Martin Moche**, thank you for taking care of the crystallization platform and for sharing my enthusiasm at BESSY when I collected my first data set. **Helena, Emma, Eleonora, and Ida** for all the good moments at the lab, fika, lunches, conference, parties…

Tomas thank you for all the great moments, for the support and help whenever needed. Your laid-back attitude and great sense of humour made my days much more enjoyable.

Mina fina vänner utanför MBB, i det verkliga livet. För all våra äventyr som får mig att tänka på annat än labbmiljö och experiment. Ni betyder otroligt mycket för mig!! Nu kan ni äntligen läsa om vad jag har försökt att förklara för er! Ett speciellt tack till **Sneha** för att ha läst igenom hela manuskriptet, för att du alltid ställer upp.

JaMiFaWi-Lindvall familjen, tack för alla mysiga stunder, goda middagar och stug-häng!

Je remercie ma chère famille en Algérie et en France pour leur soutient, particulièrement ma **Tata**.

Mina kära föräldrar, tack för all omtanke och stöd. Ni har lärt mig att jobba hårt och att satsa 100 % på det jag tycker om men även att njuta av livets sköna stunder, vi har många såna stunder framför oss!

Tack till mina underbara (små)syskon **Nadia** och **Malek** för att det alltid är så galet kul när vi träffas.

Sist men inte minst, **Hamid** du är en fantastisk människa som alltid kommer att betyda mycket för mig. Tack för att du tror på mig och för all stöd och hjälp, du är helt underbar!

References

1. Crick, F. H. C. On Protein Synthesis. *Symp. Soc. Exp. Biol. XII* 139–163 (1956).

2. Crick, F. H. C. Cendral Dogma of Moleculat Biology. *Nature* **227,** 561–563 (1970).

3. Crossley, H. Origin of the word 'protein'. *Notes and Queries* **s4-IV,** 448–449 (1869).

4. Pauling, L., Corey, R. B. & Branson, H. R. The structure of proteins; two hydrogen-bonded helical configurations of the polypeptide chain. *Proc. Natl. Acad. Sci. U. S. A.* **37,** 205–211 (1951).

5. Ali, M. H. & Imperiali, B. Protein oligomerization: How and why. *Bioorganic and Medicinal Chemistry* **13,** 5013–5020 (2005).

6. Ulmschneider, M. B., Sansom, M. S. P. & Di Nola, A. Properties of integral membrane protein structures: Derivation of an implicit membrane potential. *Proteins Struct. Funct. Genet.* **59,** 252–265 (2005).

7. Johnson, J. E. & Cornell, R. B. Amphitropic proteins: regulation by reversible membrane interactions (review). *Mol. Membr. Biol.* **16,** 217–235 (1999).

8. Silhavy, T. J., Kahne, D. & Walker, S. The bacterial cell envelope. *Cold Spring Harb. Perspect. Biol.* **2,** a000414 (2010).

9. Van Meer, G., Voelker, D. R. & Feigenson, G. W. Membrane lipids: where they are and how they behave. *Nat. Rev. Mol. Cell Biol.* **9,** 112–124 (2008).

10. Knævelsrud, H. & Simonsen, A. Lipids in autophagy : Constituents , signaling molecules and cargo with relevance to disease. *BBA - Mol. Cell Biol. Lipids* **1821,** 1133–1145 (2012).

11. Wagner, S. *et al.* Consequences of membrane protein overexpression in Escherichia coli. *Mol. Cell. Proteomics* **6,** 1527–1550 (2007).

12. Studier, F. W., Rosenberg, a. H., Dunn, J. J. & Dubendorff, J. W. Use of T7 RNA polymerase to direct expression of cloned genes. *Methods Enzymol.* **185,** 60–89 (1990).

13. Baneyx, F. & Georgiou, G. In vivo degradation of secreted fusion proteins by the Escherichia coli outer membrane In Vivo Degradation of Secreted Fusion Proteins by the Escherichia coli Outer Membrane Protease OmpT. **172,** 491–494 (1990).

14. Shineberg, B. & Zipser, D. The lon Gene and Degradation of β-Galactosidase Nonsense Fragments. *J. Bacteriol.* **116,** 1469–1471 (1973).

15. Seddon, A. M., Curnow, P. & Booth, P. J. Membrane proteins, lipids and detergents: Not just a soap opera. *Biochim. Biophys. Acta - Biomembr.* **1666,** 105–117 (2004).

16. Detergents and their uses in membrane protein science. *Anatrace*

17. Williams, R. B., Bernard, R. & Callcut, D. *Protein Chromatography Methods and Protocols*. (Springer New York Dordrecht Heidelberg London, 2011).

18. Garavito, R. M. & Ferguson-Miller, S. Detergents as Tools in Membrane Biochemistry. *J. Biol. Chem.* **276,** 32403–32406 (2001).

19. Porath, J. Immobilized metal ion affinity chromatography. *Protein Expr. Purif.* **3,** 263–281 (1992).

20. Hochuli, E., Bannwarth, W. & Döbeli, H. Genetic Approach to Facilitate Purification of Recombinant Proteins with a Novel Metal Chelate Adsorbent. *Nature* **6,** 1321–1325 (1988).

21. Healthcare, G. E. & Sciences, L. *Size Exclusion Chromatography*. (2014). at <www.gelifesciences.com/handbooks>

22. Tanford, C. Protein denaturation. *advan. Prot. Chem* **23,** 5121–282 (1968).

23. Brandts, J. F. & Lin, L. N. Study of strong to ultratight protein interactions using differential scanning calorimetry. *Biochemistry* **29,** 6927–6940 (1990).

24. Koshland, E. Application of a Theory of Enzyme Specificity To Protein. *Pnas* **44,** 98–104 (1958).

25. Senisterra, G. a *et al.* Assessing the stability of membrane proteins to detect ligand binding using differential static light scattering. *J. Biomol. Screen. Off. J. Soc. Biomol. Screen.* **15,** 314–320 (2010).

26. Kurganov, B. I. Kinetics of Protein Aggregation . Quantitative Estimation of the Chaperone Like Activity in Test Systems Based on Suppression of Protein Aggregation. *Biochem. Moscow* **67,** (2002).

27. Alexandrov, A. I., Mileni, M., Chien, E. Y. T., Hanson, M. a. & Stevens, R. C. Microscale Fluorescent Thermal Stability Assay for Membrane Proteins. *Structure* **16,** 351–359 (2008).

28. The Nobel Foundation. Wilhelm Conrad Röntgen - Biographical. *nobelprize.org*

29. Kendrew, J. C. *et al.* Structure of myoglobin: A three-dimensional Fourier synthesis at 2 A. resolution. *Nature* **185,** 422–427 (1960).

30. Kendrew, J. C. *et al.* A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* **181,** 662–666 (1958).

31. Deisenhofer, J., Epp, O., Miki, K., Huber, R. & Michel, H. Structure of the protein subunits in the photosynthetic reaction centre of Rhodopseudomonas viridis at 3□ resolution. *Nature* **318,** 618–624 (1985).

32. Wüthrich, K. The way to NMR structures of proteins. *Nat. Struct. Biol.* **8,** 923–925 (2001).

33. Lengyel, J., Hnath, E., Storms, M. & Wohlfarth, T. Towards an integrative structural biology approach: combining Cryo-TEM, X-ray crystallography, and NMR. *J. Struct. Funct. Genomics* 117–124 (2014).

34. Rhodes, G. *Crystallography Made Crystal Clear (Third Edition)*. (Academic Press).

35. Chayen, N. E. & Saridakis, E. Protein crystallization: from purified protein to diffraction-quality crystal. *Nat. Methods* **5,** 147–153 (2008).

36. Newby, Z. E. R. *et al.* A general protocol for the crystallization of membrane proteins for X-ray structural investigation. *Nat. Protoc.* **4,** 619–637 (2009).

37. Griffin, L. & Lawson, a. Antibody fragments as tools in crystallography. *Clin. Exp. Immunol.* **165,** 285–291 (2011).

38. Bukowska, M. a. & Grütter, M. G. New concepts and aids to facilitate crystallization. *Curr. Opin. Struct. Biol.* **23,** 409–416 (2013).

39. Kang, H. J., Lee, C. & Drew, D. Breaking the barriers in membrane protein crystallography. *Int. J. Biochem. Cell Biol.* **45,** 636–44 (2013).

40. Dong, A. *et al.* In situ proteolysis for protein crystallization and structure determination. *Nat. Methods* **4,** 1019–1021 (2007).

41. Sylvie, D. in *Volume 1* (ed. Walker, J. M.) (Humana Press Inc., 2007).

42. Parker, J. L. & Newstead, S. Current trends in alpha-helical membrane protein crystallization: An update. *Protein Sci.* **21,** 1358–1365 (2012).

43. Wheeler, M. J., Russi, S., Bowler, M. G. & Bowler, M. W. Measurement of the equilibrium relative humidity for common precipitant concentrations: Facilitating controlled dehydration experiments. *Acta Crystallogr. Sect. F Struct. Biol. Cryst. Commun.* **68,** 111–114 (2012).

44. Duke, E. M. H. & Johnson, L. N. Macromolecular crystallography at synchrotron radiation sources: current status and future developments. *Proc. R. Soc. A Math. Phys. Eng. Sci.* **466,** 3421–3452 (2010).

45. Garman, E. F. Developments in X-ray Crystallographic Biological Macromolecules. **921,** 134–141 (2013).

46. Hartman, P. L. Early experimental work on synchrotron radiation. *Synchrotron Radiat. News 1* **1,** (1988).

47. Brown, G., Halbach, K., Harris, J. & Winick, H. Wiggler and Undulator Magnets - a Review. *Nucl. instruments methods Phys. Res.* **208,** 65–77 (1982).

48. Bilderback, D. H., Elleaume, P. & Weckert, E. Review of third and next generation synchrotron light sources. *J. Phys. B At. Mol. Opt. Phys.* **38,** S773–S797 (2005).

49. Neutze, R. & B, P. T. R. S. Opportunities and challenges for time-resolved studies of protein structural dynamics at X-ray free-electron lasers. *Philos. Trans. R. Soc. B* **369,** (2014).

50. Liu, W. *et al.* Femtosecond crystallography of membrane proteins in the lipidic cubic phase Femtosecond crystallography of membrane proteins in the lipidic cubic phase. (2014).

51. Suga, M. *et al.* Native structure of photosystem II at 1.95 Å resolution viewed by femtosecond X-ray pulses. *Nature* (2014). doi:10.1038/nature13991

52. Moraes, I., Evans, G., Sanchez-Weatherby, J., Newstead, S. & Stewart, P. D. S. Membrane protein structure determination - the next generation. *Biochim. Biophys. Acta* **1838,** 78–87 (2014).

53. Bragg, W. H. The reflection of X-rays by crystals. *Nature* **88,** (1913).

54. Society, T. R., Society, R., Papers, C. & Character, P. The Reflexion of X-rays by Crystals. *Proc R. Soc. Lond. A* **88,** 428–38 (1913).

55. Bernhard, R. in (Taylor & Francis Group, 2009).

56. Taylor, G. The phase problem. *Acta Crystallogr. - Sect. D Biol. Crystallogr.* **59,** 1881–1890 (2003).

57. Taylor, G. L. Introduction to phasing. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66,** 325–338 (2010).

58. Crowther, R. A. Blow, D. M. A method of positioning a known molecule in an unknown crystal structure. *Acta Crystallogr.* **23,** 544–548 (1967).

59. Rossmann, M. G. The Molecular Replacement Method. *Acta Crystallogr.* **30,** 395–405 (1974).

60. Karle, J. Some development in anomalous dispersion from the structural investigation of macromolecular systems in biology. *Int. J. Quantum Chem.* **7,** 357–367 (1980).

61. Perrakis, a, Morris, R. & Lamzin, V. S. Automated protein model building combined with iterative structure refinement. *Nat. Struct. Biol.* **6,** 458–463 (1999).

62. A. L. Hodgkin, A. F. H. Propagation of Electrical Signals Along Giant Nerve Fibres. *Society* **205,** 211–30 (1979).

63. Schwiening, C. J. A brief historical perspective: Hodgkin and Huxley. *J. Physiol.* **590,** 2571–5 (2012).

64. Campagna-Slater, V. & Weaver, D. F. Molecular modelling of the GABAA ion channel protein. *J. Mol. Graph. Model.* **25,** 721–730 (2007).

65. Lai, H. C. & Jan, L. Y. The distribution and targeting of neuronal voltage-gated ion channels. *Nat. Rev. Neurosci.* **7,** 548–562 (2006).

66. Gadsby, D. C. Ion channels versus ion pumps: the principal difference, in principle. *Nat. Rev. Mol. Cell Biol.* **10,** 344–352 (2009).

67. Davidson, A. L., Dassa, E., Orelle, C. & Chen, J. Structure, function, and evolution of bacterial ATP-binding cassette systems. *Microbiol. Mol. Biol. Rev.* **72,** 317–364, table of contents (2008).

68. Jones, P. M. & George, a. M. The ABC transporter structure and mechanism: Perspectives on recent research. *Cell. Mol. Life Sci.* **61,** 682–699 (2004).

69. Goswitz, V. C. & Brooker, R. J. Structural features of the uniporter / symporter / antiporter superfamily. *Cell* 534–537 (1999).

70. Daniel, H., Spanier, B., Kottra, G. & Weitz, D. From bacteria to man: archaic proton-dependent peptide transporters at work. *Physiology (Bethesda).* **21,** 93–102 (2006).

71. Hunte, C. *et al.* Structure of a Na+/H+ antiporter and insights into mechanism of action and regulation by pH. *Nature* **435,** 1197–1202 (2005).

72. Deng, D. *et al.* Crystal structure of the human glucose transporter GLUT1. *Nature* **510,** 121–5 (2014).

73. Miller, C. ClC chloride channels viewed through a transporter lens. *Nature* **440,** 484–489 (2006).

74. Accardi, A. & Picollo, A. CLC channels and transporters: Proteins with borderline personalities. *Biochim. Biophys. Acta - Biomembr.* **1798,** 1457–1464 (2010).

75. Reddy, V. S., Shlykov, M. a., Castillo, R., Sun, E. I. & Saier, M. H. The major facilitator superfamily (MFS) revisited. *FEBS J.* **279,** 2022–2035 (2012).

76. Heymann, J. a W. *et al.* Projection structure and molecular architecture of OxIT, a bacterial membrane transporter. *Embo J.* **20,** 4408–4413 (2001).

77. Hirai, T. *et al.* Three-dimensional structure of a bacterial oxalate transporter. *Nat. Struct. Biol.* **9,** 597–600 (2002).

78. Huang, Y., Lemieux, M. J., Song, J., Auer, M. & Wang, D.-N. Structure and mechanism of the glycerol-3-phosphate transporter from Escherichia coli. *Science* **301,** 616–620 (2003).

79. Quistgaard, E. M., Löw, C., Moberg, P., Trésaugues, L. & Nordlund, P. Structural basis for substrate transport in the GLUT-homology family of monosaccharide transporters. *Nat. Struct. Mol. Biol.* **20,** 766–8 (2013).

80. Dang, S. *et al.* Structure of a fucose transporter in an outward-open conformation. *Nature* **467,** 734–738 (2010).

81. Yin, Y. Structure of the Multidrug Transporter. **741,** (2011).

82. Guettou, F. *et al.* Structural insights into substrate recognition in proton-dependent oligopeptide transporters. *EMBO Rep.* **14,** 804–10 (2013).

83. Abramson, J. *et al.* Structure and mechanism of the lactose permease of Escherichia coli. *Science* **301,** 610–615 (2003).

84. Pedersen, B. P. *et al.* Crystal structure of a eukaryotic phosphate transporter. *Nature* **496,** 533–6 (2013).

85. Shi, Y. Common folds and transport mechanisms of secondary active transporters. *Annu. Rev. Biophys.* **42,** 51–72 (2013).

86. Jardetzky, O. Simple allosteric model for membrane pumps. *Nature* **211,** 969–970 (1966).

87. Fowler, P. W. *et al.* Gating Topology of the Proton-Coupled Oligopeptide Symporters. *Structure* **23,** 290–301 (2015).

88. Smirnova, I., Kasho, V. & Kaback, H. R. Lactose permease and the alternating access mechanism. *Biochemistry* **50,** 9684–9693 (2011).

89. Zhou, Y., Guan, L., Freites, J. A. & Kaback, H. R. Opening and closing of the periplasmic gate in lactose permease. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 3774–3778 (2008).

90. Smirnova, I., Kasho, V., Sugihara, J. & Kaback, H. R. Probing of the rates of alternating access in LacY with Trp fluorescence. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 21561–21566 (2009).

91. Hvorup, R. N. & Saier, M. M. Sequence similarity between the channel-forming domains of voltage-gated ion channel proteins and the C-terminal domains of secondary carriers of the major facilitator superfamily. *Microbiology* **148,** 3755–3756 (2002).

92. Forrest, L. R., Krämer, R. & Ziegler, C. The structural basis of secondary active transport mechanisms. *Biochim. Biophys. Acta* **1807,** 167–188 (2011).

93. Krishnamurthy, H., Piscitelli, C. L. & Gouaux, E. Unlocking the molecular secrets of sodium-coupled transporters. *Nature* **459,** 347–355 (2009).

94. Harder, D. *et al.* DtpB (YhiP) and DtpA (TppB, YdgR) are prototypical proton-dependent peptide transporters of Escherichia coli. *FEBS J.* **275,** 3290–3298 (2008).

95. Terada, T. & Inui, K. I. *Recent advances in structural biology of peptide transporters*. *Current Topics in Membranes* **70,** (Elsevier, 2012).

96. Biegel, a. *et al.* The renal type H+/peptide symporter PEPT2: Structure-affinity relationships. *Amino Acids* **31,** 137–156 (2006).

97. Stone, H. L. & Holmes, K. R. Amarican physiological society fall meeting. *Am. Physiol. Soc.* **13,** (1970).

98. Rubio-Aliaga, I. & Daniel, H. Peptide transporters and their roles in physiological processes and drug disposition. *Xenobiotica.* **38,** 1022–1042 (2008).

99. Meredith, D. & Price, R. a. Molecular modeling of PepT1 - Towards a structure. *J. Membr. Biol.* **213,** 79–88 (2006).

100. Newstead, S. *et al.* Crystal structure of a prokaryotic homologue of the mammalian oligopeptide-proton symporters, PepT1 and PepT2. *EMBO J.* **30,** 417–426 (2011).

101. Doki, S. *et al.* Structural basis for dynamic mechanism of proton-coupled symport by the peptide transporter POT. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 11343–8 (2013).

102. Solcan, N. *et al.* Alternating access mechanism in the POT family of oligopeptide transporters. **31,** 3411–3421 (2012).

103. Lyons, J. a *et al.* Structural basis for polyspecificity in the POT family of proton-coupled oligopeptide transporters. *EMBO Rep.* **15,** 886–93 (2014).

104. Doki, S. *et al.* Structural basis for dynamic mechanism of proton-coupled symport by the peptide transporter POT. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 11343–8 (2013).

105. Zhao, Y. *et al.* Crystal structure of the E. coli peptide transporter YbgH. *Structure* **22,** 1152–1160 (2014).

106. Bork, P. & Koonin, E. Protein sequence motifs. *Curr. Opin. Struct. Biol.* 366–376 (1996).

107. Jiang, D. *et al.* Structure of the YajR transporter suggests a transport mechanism based on the conserved motif A. *Proc. Natl. Acad. Sci. U. S. A.* **110,** 14664–9 (2013).

108. Ito, K. *et al.* Analysing the substrate multispecificity of a proton-coupled oligopeptide transporter using a dipeptide library. *Nat. Commun.* **4,** 2502 (2013).

109. Guettou, F. *et al.* Selectivity mechanism of a bacterial homolog of the human drug-peptide transporters PepT1 and PepT2. *Nat. Struct. \& Mol. Biol.* **21,** 728–731 (2014).

110. Mitchell, I. C. W. and P. Stoicheiometry of Lactose-Protein Symport Across the Plasma Membrane of Escherichia coil. *Methods* 587–592 (1973).

111. Smirnova, I., Kasho, V., Sugihara, J., Vazquez-Ibar, J. L. & Kaback, H. R. Role of protons in sugar binding to LacY. *Proc. Natl. Acad. Sci.* **109,** 16835–16840 (2012).

112. Ethayathulla, A. S. *et al.* Structure-based mechanism for Na(+)/melibiose symport by MelB. *Nat. Commun.* **5,** 3009 (2014).

113. Parker, J. L., Mindell, J. a & Newstead, S. Thermodynamic evidence for a dual transport mechanism in a POT peptide transporter. 1–13 (2014). doi:10.7554/eLife.04273

114. Lyons, J. a *et al.* Structural basis for polyspecificity in the POT family of proton-coupled oligopeptide transporters. *EMBO Rep.* **15,** 886–93 (2014).

115. Miroux, B. & Walker, J. E. Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels. *J. Mol. Biol.* **260,** 289–298 (1996).

116. Frauenfeld F, Löving R, Zhang Y, Zhu L, Jegerschöld C, Guettou F, Moberg P, Löw C, Nyström N, G. H. and N. P. A multi-functional nanoparticle system based on a small human protein. *Nat. Nanotechnol*

117. Bruhn, H. A short guided tour through functional and structural features of saposin-like proteins. *Biochem. J.* **389,** 249–257 (2005).